DOCUMENT RESUME

ED 083 779 EC 060 410

TITLE	Studies in Achievement Testing, Hearing Impaired Students. United States: Spring 1971.
INSTITUTION	Gallaudet Coll., Washington, D.C. Office of Demographic Studies.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Jul 73
NOTE	86p.; Annual survey of hearing impaired children and youth
AVAILABLE FROM	Gallaudet College Book Store, Washington, D.C. 20002 ($2.00)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; Achievement Tests; *Aurally Handicapped; *Exceptional Child Research; Performance Factors; Test Construction; Testing; Test Interpretation; *Test Reliability; *Test Validity
IDENTIFIERS	*Stanford Achievement Test

ABSTRACT

Results of three studies based on Stanford Achievement Test data collected during the National Achievement Testing Program for Hearing Impaired Students in Spring, 1971 are reported. Compared in the first study are performance patterns of hearing impaired students on the Intermediate I and Advanced batteries with patterns of hearing standardization groups on the same test levels. Content sub-areas within the subtests are analyzed and areas in which hearing impaired students performed better or worse than the standardization group are identified. Technical problems relating to test construction are noted. In the second study, the discriminative validity of items in five selected subtests of the Intermediate I battery is examined and whether the items effectively distinguished between high and low achieving students is determined. Examples of items for each of the five subtests are presented to show high and low discrimination coefficients. The paragraph meaning and science subtests are said to have the highest proportions of low discriminating items and thus the lowest average discrimination coefficients. Reported last are findings of a reliability study of the Primary II battery which involved use of Alternate Forms X and W for the hearing impaired with a sample of 192 students. The standard errors of measurement and other related data are presented. Included in appendixes are descriptions and limitations of the achievement testing program, testing procedures, and a list of schools and classes that participated in the program. (MC)

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED 083779

EC060410

This research was supported by a grant from the National Institute of Education, U.S. Department of Health, Education & Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

OFFICE OF DEMOGRAPHIC STUDIES
Publication – Series D – Number 11

For Sale by Gallaudet College Book Store
Washington, D.C. 20002 – Price Two Dollars

LIBRARY OF CONGRESS CATALOG CARD NUMBER 73-85860

ED 083779

SERIES D
NUMBER 11

STUDIES IN ACHIEVEMENT TESTING, HEARING IMPAIRED STUDENTS

UNITED STATES: SPRING 1971

DATA FROM THE ANNUAL SURVEY OF HEARING IMPAIRED CHILDREN AND YOUTH

OFFICE OF DEMOGRAPHIC STUDIES
GALLAUDET COLLEGE

Washington, D.C.

July 1973

GALLAUDET COLLEGE

Edward C. Merrill, Jr., Ph.D., President
John S. Schuchman, Ph.D., Dean

OFFICE OF DEMOGRAPHIC STUDIES
Peter Ries, Acting Director

SENIOR STAFF MEMBERS
Neil Murphy, Statistical Analyst
Brenda Rawlings, Statistical Analyst
Arthur Schildroth, Statistical Analyst
Ray Trybus, Ph.D., Research Psychologist
Sally Wagner, Administrative Assistant

TABLE OF CONTENTS

PAGE

Acknowledgements v

Abstract vi

General Introduction 1

"Patterns of Achievement Test Performance" 3

Detailed Tables of Achievement Test Results

Intermediate I Battery 21

Advanced Battery 33

"The Discriminative Validity of Selected Sub-tests of The Stanford Achievement Test, Intermediate I Battery, for Hearing Impaired Students: Spring 1971" 47

"The Reliability of the Stanford Achievement Test, Primary II Battery, for Hearing Impaired Students" 63

Appendices

Appendix I: Description, Qualifications, and Limitations of the Achievement Testing Program 69

Appendix II: Standardized Testing Procedures Developed for the Spring 1971 Achievement Testing Program 73

Appendix III: Schools and Classes That Participated in the Achievement Testing Program 75

ACKNOWLEDGEMENTS

The Office of Demographic Studies wishes to thank those educational programs that participated in the Achievement Testing Program. The names of these programs are listed in Appendix III. In addition we would like to extend our appreciation to those schools which assisted in the special reliability study that was undertaken. The schools and staff that provided this support are listed below.

Arizona State School for the Deaf and the Blind, Tucson, Arizona
- June Newkirk
- Earl Lubbers
- William Tellam

The Austine School for Hearing Handicapped Children, Brattleboro, Vermont
- Raymond Stevens
- Robert Brenton
- Elliott Engleheart

Houston Independent School District, Houston, Texas
- Edna Mae Vaughan
- Vassar Dickerson
- Katy Thierfelder
- Dorothy Priddy

Kendall School, Washington, D.C.
- J. Leland Clack
- Pamela Minger
- Maryann McDermott

Litzsinger School, St. Louis, Missouri
- Margot Gossin
- Barbara Braznell
- Sue Freeman

The Maryland School for the Deaf, Frederick, Maryland
- Kenneth Kritz
- Gertrude Galloway
- Linda Rudy

Seattle Public Schools, Seattle, Washington
- Delbert Royer
- Connie Jorgenson
- Linda Abbott

The Virginia School for the Deaf and the Blind, Staunton, Virginia
- John Shipman
- Milton Yoder
- Earnest Guffey

Recognition is also given to Mr. Richard Meisegeier of Gallaudet College who trained the test observers for the reliability study.

As always the support and assistance of our National Advisory Committee has been most helpful in guiding this achievement testing project and other activities of the Annual Survey.

National Advisory Committee

John D. Harrington, Ed.D., Chairman
Richard G. Brill, Ph.D., Vice-Chairman

Ellery J. Adams
H. W. Barkuloo, Ph.D.
Bill G. Blevins
Wallace T. Bruce
Edward C. Carney
Helen B. Craig, Ph.D.
Robert T. Dawson
C. Joseph Giangreco, Ph.D.
Lloyd Harrison
Marshall S. Hester, Litt.D.
William McConnell
Sister Rita McCormack
Noel Matkin, Ph.D.
Thomas A. Mayes, Ph.D.
John Melcher
Frank Powell
Frederick C. Schreiber
Jess M. Smith
Roy M. Stelle, Litt.D.
Armin G. Turechek, Ed.D.
Gerard G. Walter, Ed.D.

Sal DiFrancesca, Ph.D.
Raymond Trybus, Ph.D.
Carol Buchanan
Washington, D.C.
July, 1973

ABSTRACT

This publication presents three studies based on Stanford Achievement Test data collected during the National Achievement Testing Program for Hearing Impaired Students conducted by the Office of Demographic Studies in spring, 1971.

The first study compares the patterns of performance of hearing impaired students on the Intermediate I and Advanced batteries to the patterns of the hearing standardization groups for those test levels. The analysis concentrates primarily on content sub-areas within the sub-tests (e.g., addition, subtraction, etc., within Arithmetic Computation) and identifies those areas in which hearing impaired students performed bc ter then the standardization group, and vice versa. Some technical problems relating to test construction are also presented.

The second study examines the discriminative validity of the items in five selected sub-tests of the Intermediate I battery. Since the purpose of achievement tests is to differentiate levels of achievement within groups of students, it is essential that the items be able to distinguish effectively between high achieving and low achieving students. This study reviews the evidence in this respect and presents examples for each of the five sub-tests of items exhibiting especially high and especially low discrimination coefficients. The average discrimination coefficients ranged from .32 to .61 for the five sub-tests. The Paragraph Meaning and Science Sub-tests had the highest proportions of low discriminating items and thus the lowest average discrimination coefficients.

The final study reports the findings of a reliability study of the Primary II battery which was conducted in spring, 1971, in conjunction with the National Testing Program. Alternate Forms X and W-HI of the Primary II battery were administered to 192 students in a specially selected national sample with an interval of two to four weeks between the first and second testing sessions. The standard errors of measurement and other related data are also presented.

We should like to express our appreciation to the publishers of the Stanford Test Series for their permission to use the following items, which appear on the designated pages of this publication:

Items from the Intermediate I battery:
- Paragraph Meaning, Nos. 2, 19, 53, 5, on page 56
- Spelling, all items, on pages 23, 57
- Arithmetic Computation, Nos. 1, 9, 26, on pages 57, 58
- Arithmetic Applications, Nos. 11, 18, 21, 26, on pages 58, 59
- Science, Nos. 17, 36, 56, on page 60

Items from the Advanced battery:
- Spelling, all items, on page 34

Table from the Technical Supplement, Table 10, on page 70.

These items and table were reproduced from Stanford Achievement Test, copyright © 1964-1966 by Harcourt Brace Jovanovich, Inc. They are reproduced by permission.

Studies in Achievement Testing, Hearing Impaired Students United States: Spring 1971

INTRODUCTION

Many individuals concerned with the education of the hearing impaired have indicated their interest in the measurement of the academic achievement of hearing impaired students. The need for national data on the achievement levels of hearing impaired students together with a deliberate research effort to determine the appropriateness and suitability of standard achievement tests for this student population has been felt by many of those who bear the responsibility of educating the hearing impaired.

In response to this need, the Annual Survey of Hearing Impaired Children and Youth has devoted part of its resources to collecting and analyzing achievement testing information on students attending special educational programs for the hearing impaired. The longer range purposes of this activity are to determine the suitability of existing achievement tests for these students and to develop procedures and materials designed to enhance the validity and reliability of achievement testing results.

This publication presents the results of three studies undertaken with the data collected on a national group of hearing impaired students who were administered the Stanford Achievement Test in the spring of 1971. The demographic characteristics of the nearly 17,000 hearing impaired students who participated in the testing program are presented in Appendix III of a previous Annual Survey publication, Series D-9. Also shown in that Appendix are the distributions of these characteristics for the 41,109 hearing impaired students that were reported to the Annual Survey of Hearing Impaired Children and Youth for the 1970-71 school year.

The first study is concerned with similarities and dissimilarities in performance patterns of the hearing impaired students in comparison to the normal hearing students on which the test was standardized. The percentages of each type of student who correctly answered each of the items in all of the sub-tests [except the Word Study Skills] of the Intermediate I and Advanced batteries are presented along with a discussion of the highlights of these data. While the comparison is not controlled for age and the hearing impaired students were in general four or five years older than the hearing students, identification of the types of items on which each group showed superior performance should be of interest, and hopefully of use, to those interested in determining areas of relative weaknesses and strengths in the academic performance of hearing impaired students.

The third study deals with an investigation of the reliability of a traditional academic achievement test when used with students in special educational programs for the hearing impaired. Alternative forms of the Primary II battery were administered to a national sample of 178 hearing impaired students. The results indicate reliability coefficients for the various sub-tests that range from adequate to high.

The question of validity is, of course, far more complex than that of reliability. The second study deals with this question only in the limited sense of considering the ability of the items in some of the sub-tests of the Intermediate I battery to discriminate between students who receive high and low scores on the particular sub-tests. The results of this study indicate that a significant proportion of the items examined do discriminate well between the two types

of students. Also, specific items that do and do not discriminate are identified and discussed.

Two previous publications on the results of the National Achievement Testing Program conducted by the Annual Survey of Hearing Impaired Children and Youth have already appeared. In each of these the background of this testing program and the many qualifications relating to use of its results have appeared at the beginning of the publications. On the basis of an assumption that those interested in the results of this testing program have already twice read this material, this information appears as Appendix I of this publication. However, anyone not familiar with this material, especially the qualifications of the data upon which these studies are based, should read Appendix I. Appendix II will also provide background information on the standardized testing procedures developed for the 1971 Achievement Testing Program.

Patterns of Achievement Test Performance

Raymond J. Trybus, Ph.D., and Carol Buchanan

INTRODUCTION

Purpose of this Study

That hearing impaired students do not perform in the same manner as normally hearing students on the Stanford Achievement Test is a fact well known to educators of hearing impaired students and extensively documented in the reports of the two National Achievement Testing Programs conducted by the Annual Survey of Hearing Impaired Children and Youth in 1969 and 1971. Publications D-1, D-2, D-8, and D-9 from the Annual Survey[1] provide extensive data which permit various comparisons on the basis of whole sub-tests among various age groups, groups with varying degrees of hearing loss, and between hearing impaired students and the normally hearing standardization group on which the Stanford scores are based.

The purpose of the present study is to extend this investigation of the test performance of hearing impaired students by making detailed comparisons of the relative performance of the two groups on the Intermediate I and Advanced level batteries. The items which comprise the various sub-tests were broken down into subgroupings on the basis of specific content matter or other relevant dimensions, and the performance of the two groups was compared in order to elucidate areas of specific strength and weakness. The potential value of this information for curriculum planning is apparent, with the caution that the data here were gathered on a national basis and thus may or may not reflect accurately the situation at a particular school. Similar analyses of local data, in themselves and by comparison with these national figures, can yield much that is of value to educators and curriculum designers. The examination of "response position" effects may be of less immediate relevance for curriculum planning in the usual subject matter areas, but it presents a technical measurement problem which needs to be handled and could conceivably form the basis of instruction on problem-solving skills and, more mundanely, on test-taking procedures themselves.

Qualifications and Limitations of the Data

The statements of qualification and limitation which appear in Appendix I to this publication apply to these data, as they do to all data based on the Annual Survey's two National Achievement Testing Programs.

In addition to these general statements, some further qualifications which apply specifically to this study of performance patterns need to be described here. This study is based on the Intermediate I and Advanced levels, Form W, of the 1964 edition of the Stanford Achievement Test. Whether this information can be generalized to the other levels of the Stanford is unknown. While the format and "testing philosophy" of the other levels are very similar, the content is different, and it is on the specifics of content that this study is focusing.

The specific characteristics of the students who took these two test levels will now be considered. Table 1 summarizes the available information on the

[1] Gallaudet College, Office of Demographic Studies. *Academic Achievement Test Performance of Hearing Impaired Students—United States: Spring 1969.* Series D, Number 1.

Gallaudet College, Office of Demographic Studies. *Item Analysis of Academic Achievement Tests: Hearing Impaired Students—United States: Spring 1969.* Series D, Number 2.

Gallaudet College, Office of Demographic Studies. *Item Analysis of an Achievement Testing Program for Hearing Impaired Students—United States: Spring 1971.* Series D, Number 8.

Gallaudet College, Office of Demographic Studies. *Academic Achievement Test Results of a National Testing Program for Hearing Impaired Students—United States: Spring 1971.* Series D, Number 9.

TABLE 1: CHARACTERISTICS OF STUDENTS TAKING THE INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
A: Better Ear Averages (ISO)	**Number of Students**	**Percent**	**Number of Students**	**Percent**
Total Students*	2,398	100.0	654	100.0
59dB & Below	267	11.1	85	13.0
60 to 98 dB	1,414	59.0	390	59.6
99dB & Above	717	29.9	179	27.4
B: Age	**Hearing**	**Hearing Impaired**	**Hearing**	**Hearing Impaired**
Total Students	618	2,821	703	727
Range of Years	9 to 12	6 to 21+**	12 to 15	12 to 21+**
Mean Age in Years	9.9	15.4	12.9	16.9
Standard Deviation Age (Years)	0.5	2.5	0.6	1.8

*Excluded are those for whom a better ear average could not be computed and those for whom no audiological data were reported.

**The age ranges reported for the hearing impaired students exclude one case at the Intermediate I level and four cases at the Advanced level which are highly improbable and may reflect errors in scoring or in the ages reported for the students who took the tests.

hearing loss levels of the hearing impaired students and on the ages of the hearing impaired students and the standardization group students.

The numbers of subjects reported for the hearing groups are those included in the equating of forms program,[1] from which the item difficulty figures for Form W used in this report were obtained. The age ranges, means, and standard deviations reported for the hearing groups are estimates based on the age distributions of the entire standardization population. The item difficulty figures for the standardization group are based on Grade 4 students for the Intermediate I level, and on Grade 7 students for the Advanced level; the age ranges, means, and standard deviations are estimates for students in these grade levels.

In interpreting the data in this report, it is important to keep these characteristics in mind. For example, while the data indicate that the hearing impaired students showed better performance than the standardization group in some areas, it must be understood that the hearing impaired students who took the Intermediate I were, on the average, 5.5 years older than their hearing counterparts. The hearing impaired students who took the Advanced battery were, on the average, 4.0 years older than their counterparts in the standardization group.

The distributions of degree of hearing loss presented in Table 1-A are clearly similar and are not significantly different upon statistical test ($X^2 = 2.84$, 2 d.f., $p > .20$). Observed differences in performance between the two hearing impaired groups cannot, therefore, be attributed merely to differences in the degree of hearing loss.

OVERALL PATTERN SIMILARITIES AND DIFFERENCES

The basic statistic which will be used throughout this study is the item analysis or item difficulty index. This is simply the percentage of individuals in a group who gave the correct answer to the item in question. An easy item is answered correctly by many students and the item difficulty figure is, for example, 85 percent. Another more difficult item is answered correctly by few students and the difficulty figure might be, for example, 27 percent. Since the higher number indicates *less* difficulty, it might be more appropriate to call this an item *easiness* index, but the more traditional usage will be retained here, i.e., "item difficulty."

[1]The equating of forms program was conducted by Harcourt Brace Jovanovich, Inc., publishers of the Stanford tests, to determine the psychometric equivalence of Forms W, X, Y, Z, and S of the 1964 edition of the Stanford Achievement Test Series.

The first comparison to be made is based on the average item difficulty index for each sub-test. This is obtained by taking the arithmetic mean of the item difficulty indexes for all the items in a given sub-test or part of a sub-test. Table 2 gives these mean figures for both the Intermediate I and the Advanced levels and for both the hearing impaired group and the standardization group.[1] The differences noted here are generally quite small, reflecting the overall adequacy of these tests for hearing impaired students. Because of the small differences, however, these mean figures contribute little to a detailed understanding of the differences between the two groups.

Some general index of the similarity of the performance *patterns* of the two groups would prove useful. Such an index was devised in the following manner. For each item in these batteries there exists a pair of item difficulty values, one for the hearing impaired group, the other for the standardization group. By correlating the corresponding item difficulty values for the two groups across all the items of a sub-test, an index is obtained which indicates the degree of similarity of the difficulty patterns on that sub-test for the two groups. If both groups tended to find the same items easy and the same items difficult, the items would be ranked in about the same order of difficulty for the two groups, and the index would be high, for example, .89. As the difficulty patterns diverge more and more, the index of similarity would decrease correspondingly. Like all correlations, this statistic has a theoretical range from +1.00 to −1.00. The similarity index values for all sub-tests of the Intermediate I and Advanced batteries are given in Table 3.

Inspection of this table shows that the similar-

[1] The standardization group item analysis figures are those for Form W as supplied by the test publisher and are based on figures obtained in the "equating of forms" program conducted for the 1964 edition.

TABLE 2: MEAN ITEM DIFFICULTY FIGURES FOR SUB-TESTS AND PARTS OF SUB-TESTS, INTERMEDIATE I AND ADVANCED LEVELS, STANFORD ACHIEVEMENT TEST, 1964 EDITION.

	Intermediate I			Advanced		
		Mean Item Difficulty Index			Mean Item Difficulty Index	
Sub-test	Number of Items	Standardization Group	Hearing Impaired Group	Number of Items	Standardization Group	Hearing Impaired Group
Battery Total	479*	58*	53*	532	56	54
Word Meaning	38	55	38	**	**	**
Paragraph Meaning	60	53	40	60	55	49
Spelling	50	62	73	58	52	58
Language (Total)	122	64	57	145	67	65
Usage	38	61	48	38	51	49
Punctuation	20	54	53	20	59	61
Capitalization	36	80	79	45	87	87
Dictionary Use	10	53	35	24	60	52
Sentence Sense	18	60	50	18	70	64
Arithmetic Computation	39	59	69	41	52	55
Arithmetic Concepts	32	57	50	40	49	49
Arithmetic Applications	33	54	46	36	43	39
Social Studies (Total)	49	54	53	92	52	47
Content	24	52	44	52	50	44
Study Skills	25	56	60	40	55	52
Science	56	56	46	60	52	50

*Figures do not include Word Study Skills Sub-test.

**This sub-test is not included in the Advanced battery.

TABLE 3: INDEX OF SIMILARITY OF DIFFICULTY PATTERNS FOR THE INTERMEDIATE I AND ADVANCED BATTERIES, FORM W, STANFORD ACHIEVEMENT TEST, BY SUB-TESTS AND MAJOR SUBDIVISIONS WITHIN SUB-TESTS.

Sub-test	Intermediate I		Advanced	
	Correlations Between Hearing Impaired & Standardization Group Item Difficulty Percents	Number of Items	Correlations Between Hearing Impaired & Standardization Group Item Difficulty Percents	Number of Items
Battery Total	.70	540	.84	532
Word Meaning	.81	38	**	**
Paragraph Meaning	.66	60	.81	60
Spelling	.91	50	.93	58
Word Study Skills Total	.65	61	**	**
A. Phonics	.60	36	**	**
B. Syllabication	.81	25	**	**
Language Total	.7	122	.86	145
A. Usage	.68	38	.80	38
B. Punctuation	.80	20	.76	20
C. Capitalization	.90	36	.90	45
D. Dictionary Skills	.63	10	.70	24
E. Sentence Sense	.27	18	.30	18
Arithmetic Computation	.93	39	.81	41
Arithmetic Concepts	.68	32	.65	40
Arithmetic Applications	.75	33	.92	36
Social Studies Total	.83	49	.79	92
A. Content	.85	24	.83	52
B. Study Skills	.85	25	.74	40
Science	.80	56	.89	60

**This sub-test not included in the Advanced battery.

ity indices vary from .27 on Sentence Sense to .93 on Arithmetic Computation in the Intermediate I battery. In the Advanced battery the figures range from .30 for Sentence Sense to .93 on Spelling. In other words, the Intermediate I Arithmetic Computation and the Advanced Spelling present very much the same challenge to the hearing impaired group that they presented to the standardization group. Sentence Sense, however, is a very different task for the two groups; items that were difficult for the standardization group are easier for the hearing impaired group, and items that were easy for the standardization group proved more difficult for the hearing impaired group. Understanding of these similarities and differences will depend on the detailed examination of the sub-tests' content and format in the following sections.

In a final look at the data from an overall battery point of view, the items in each sub-test were divided into three groups. The first group consisted of items on which the standardization group had a higher item difficulty index (i.e., a greater proportion of hearing students answered the item correctly) than the hearing impaired group. The second group contained items on which there was no difference in item difficulty or on which the differences were trivial. The third group consisted of those items on which the hearing impaired group had a higher difficulty index than the standardization group.

Whether or not a difference was trivial was determined as follows. The statistical test for the significance of the difference between two proportions allows the determination of whether two given proportions are significantly different from each other. In using this test, one enters the two given proportions into the formula and determines from the resulting figure whether the difference is significant at a predetermined p-level. For purposes of this study, the final figure required for a difference significant at the .05 level was supplied, and the

formula was solved for the required amount of difference between the two proportions. Since the results depend on the absolute magnitude of the two proportions, repeated solutions were required. The results of this procedure are given in Table 4 below. Taking the average of the two proportions involved (using percentage notation), one enters the appropriate row of the table and locates the minimum difference in percentage points required for statistical significance. Items which displayed differences smaller than those listed constituted the second, "no difference" group.

The results of this classification procedure are shown in Table 5, which shows the percentage of items in a sub-test falling into each of these three groups for all sub-tests of the Intermediate I and Advanced batteries. In making use of these figures it is necessary to recall the caution stated in the Introduction, i.e., on the average, hearing impaired students taking the Intermediate I sub-test are 5.5 years older than their standardization group counterparts. Hearing impaired students receiving the Advanced battery are, on the average, 4.0 years older than their counterparts in the standardization group.

CONTENT AND FORMAT ANALYSES OF THE SUB-TESTS

The following sections of this report will focus on analyses of parts, sections, and item groups within individual sub-tests in order to elucidate further the actual differences between the hearing impaired and the standardization group. In each case, corresponding sub-tests of the Intermediate I and the Advanced batteries will be considered together, although content differences do not always permit presentation of the data from both levels in a single table.

The analyses presented in the following sections represent the "highlights" of the data available on these two test levels. For the reader who wishes to follow the discussion by referring to the specific items involved, and for the reader who wishes to examine the data for himself and perhaps construct other item groups for comparison, the basic data on which these "highlights" were based are presented in Detailed Tables on pages 21 through 44. Tables Int-A through Int-N present data from the Intermediate I battery. Tables Adv-A through Adv-M present data from the Advanced battery. In each case, the table lists the item number (its position in the test booklet), the content categorization or specification, the item difficulty value for hearing impaired students, the item difficulty value for the standardization group of normally hearing students, and the item difficulty difference figure, defined as the hearing impaired group figure minus the standardization group figure.

Word Meaning

This sub-test is not included in the Advanced battery, and thus this discussion is confined to the Intermediate I level. It is a difficult test for the hearing impaired students, in that they perform less well than the standardization group on 79 percent of the items (see Table 5). The detailed table for this

TABLE 4: DIFFERENCES IN ITEM DIFFICULTY INDEXES REQUIRED FOR DETERMINATION OF "NO DIFFERENCE" ITEMS.

Battery Level	Average of Two Item Difficulty Figures for the Two Groups on Single Item	Difference Required for Statistical Significance
Intermediate I	5% to 15%	3%
	16% to 35%	4%
	36% to 65%	5%
	66% to 85%	4%
	86% to 95%	3%
Advanced	5% to 15%	3%
	16% to 25%	4%
	26% to 75%	5%
	76% to 85%	4%
	86% to 95%	3%

TABLE 5: PERCENTAGE OF ITEMS IN THREE PERFORMANCE CATEGORIES FOR ALL SUB-TESTS OF THE INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I			Advanced		
Sub-test	Standardization Group Better	No Difference	Hearing Impaired Group Better	Standardization Group Better	No Difference	Hearing Impaired Group Better
Battery Total	51.1	19.5	29.4	38.6	28.9	32.5
Word Meaning	79	13	8	**	**	**
Paragraph Meaning	60	27	13	57	25	18
Spelling	2	14	84	0	19	81
Word Study Skills Total	89	10	1	**	**	**
A. Phonics	92	6	2	**	**	**
B. Syllabication	84	16	0	**	**	**
Language Total	48	23	29	40	30	30
A. Usage	74	5	21	40	34	26
B. Punctuation	30	25	45	35	10	55
C. Capitalization	19	42	39	24	45	31
D. Dictionary Skills	80	20	0	71	12	17
E. Sentence Sense	56	22	22	45	33	22
Arithmetic Computation	5	18	77	27	27	45
Arithmetic Concepts	50	31	19	40	22	38
Arithmetic Applications	64	18	18	53	28	19
Social Studies Total	41	20	39	51	29	20
A. Content	54	29	17	60	27	13
B. Study Skills	28	12	60	40	32	28
Science	66	18	16	33	45	22

**This sub-test not included in the Advanced battery.

sub-test is Table Int-A on page 21.

Categorization of the items by part of speech involved is difficult because of the format of the items. The categories could refer to the word or phrase in the stem, to the correct response choice, or to the set of all response choices. The method used in this study was to categorize on the basis of the correct response option. Table 6 presents the results obtained.

TABLE 6: CONTENT ANALYSIS OF THE WORD MEANING SUB-TEST OF THE INTERMEDIATE I BATTERY.

Part of Speech	Number of Items	Average Difference
Noun	11	−13.9
Verb	12	−13.0
Adjective or Adverb	15	−20.9

The "Average Difference" in the table refers to the difference between the item difficulty figures for the two groups. It is figured by taking the hearing impaired group figure, minus the standardization group figure. A negative result indicates items on which hearing impaired students performed less well than the standardization group. Positive figures indicate items on which the hearing impaired group showed better performance, or higher item difficulty figures. It is apparent from Table 6 that the hearing impaired students performed less well in all categories, but that the adjective and adverb items were especially difficult, relative to their difficulty for the standardization group.

Paragraph Meaning

This test, intended to measure understanding of connected written language, consists of 60 multiple choice items at each level. Two types of items are included. The first involves supplying a missing word

TABLE 7: ITEM FORMAT ANALYSIS OF THE PARAGRAPH MEANING SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
Type of Item	Number of Items	Average Difference	Number of Items	Average Difference
Type I—Missing Word(s)	50	−15.7	45	− 4.5
Type II—Question or Completion	10	+ 0.5	15	−11.2

or words within the text of the passage to be read; these items predominate at both levels. The second type of item involves answering direct questions or completing a statement which follows the reading passage, the answer to which is contained or implied in the preceding passage. Table 7 compares the performance of hearing impaired students on the two types of items and on both levels of the test. The Detailed Tables for this sub-test are Tables Int-B and Adv-A on pages 22 and 33, for the Intermediate I and Advanced levels, respectively.

There is thus a reversal from one level to the other. At the Intermediate I level the hearing impaired group performed slightly better than the standardization group on Type II items and fell substantially behind on Type I items. At the Advanced level, by comparison, hearing impaired students performed more like the standardization group on Type I items and fell substantially behind on Type II items.

The items in this test can be grouped into three categories on the basis of differential depth of understanding of the printed passage required in order to select the correct response. The first category, Comprehension–Literal, requires only literal understanding of material actually present in the passage. The second, Comprehension–Inferential, requires some inferential processing of materials in the passage; and the third, Inference, requires greater understanding of the materials as well as reference to a general fund of knowledge and logic independent of the printed materials.

The items in these three categories, however, are not evenly dispersed throughout the test. The majority of the Comprehension–Literal items occur near the beginning of the test, with most of the Inference items being located near the end of the test. This position effect must therefore be taken into account in evaluating student performance on the three types of items. Table 8 shows the effects of item type and item position on the relative performance of hearing impaired students.

It is apparent from these figures that position has a greater effect on the scores than does item type.

TABLE 8: CONTENT AND POSITION ANALYSIS OF THE PARAGRAPH MEANING SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
Variable	Number of Items	Average Difference	Number of Items	Average Difference
Content Category:				
Comprehension—Literal	20	−12.7	16	−10.8
Comprehension—Inferential	28	−12.9	25	− 6.4
Inference	12	−13.8	19	− 2.0
Position in Test:				
First Third (Positions 1-20)	20	−27.9	20	−11.8
Second Third (Positions 21-40)	20	−11.2	20	− 6.3
Last Third (Positions 41-60)	20	+ 0.05	20	− 0.6

Analysis of variance indicates that at the Intermediate I level item type accounts for approximately 0.? percent of the variance, while position accounts for 85.1 percent of the variance. The corresponding figures for the Advanced level are 31.0 percent for item type, and 56.9 percent for position. At the Intermediate I level, the effects of item type are not statistically significant while the effects of position are significant well beyond the .005 level. The interaction between the two variables is significant beyond the .05 level, reflecting the fact that the differences among the three item types become smaller as the item positions move closer to the end of the test. At the Advanced level the effects of item type are significant beyond the .05 level, and the effects of position are significant beyond the .005 level. The interaction is not significant in this case.

When the difficulty difference figures (i.e., hearing impaired figure minus standardization group figure) are correlated with the length of the reading passage for that item, defined as the number of lines in the passage as it appears in the test booklet, the resulting figures indicate the degree to which the hearing impaired group's standing relative to the standardization group is a function of the passage length. In the Intermediate I battery this correlation is + .46 ($p < .01$), meaning that the longer the passage is, the more favorably the hearing impaired group compares to the standardization group. For the Advanced battery, however, the relationship is reversed, with a correlation of − .26 ($p < .05$). In this battery, the *shorter* the reading passage, the more favorably the hearing impaired group performed. In the absence of similar information for the other batteries, it is difficult to determine the significance of this finding.

Spelling

This test consists of 4-option multiple choice items, in which the test-taker is to identify which of the four words is spelled incorrectly. The Intermediate I test has 50 such items; the Advanced level has 58. The overall performance of the hearing impaired group was superior on this test, but the age difference reported above (Table 1, p. 4) must be kept in mind. The Detailed Tables for this sub-test are Tables Int-C and Adv-B on pages 23 and 34, for the Intermediate I and Advanced levels, respectively. Analysis of the results by the part of speech of the misspelled word, shown in Table 9, reveals only small variations among the categories.

Word Study Skills

This test, which appears at the Intermediate I level but not at the Advanced, consists of two parts, phonics and syllabication. Since both parts of the test depend directly on familiarity with and use of hearing, it is inappropriate for hearing impaired students, and its administration was not recommended in the National Testing Program. The resulting data for those students who did take this test[1] indicate that the scores are largely a function of the degree of hearing loss. Hearing impaired students, predictably, performed less well than the standardization group on 33 of the 36 phonics items and on 22 of the 25 syllabication items. The hearing impaired group showed better performance on only one item, a finding that can be regarded as negligible and perhaps unreliable.

[1]Gallaudet College, Office of Demographic Studies. *Academic Achievement Test Results of a National Testing Program for Hearing Impaired Students—United States: Spring 1971*. Series D, Number 9.

TABLE 9: CONTENT ANALYSIS OF THE SPELLING SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
Part of Speech	Number of Items	Average Difference	Number of Items	Average Difference
Noun	16	+12.3	28	+10.3
Verb	18	+ 9.8	14	+ 9.3
Adjective or Adverb	14	+12.4	15	+11.2
Other	2	+ 5.0	1	+14.0

Language

This sub-test consists, at both levels, of five separately administered parts: Usage, Punctuation, Capitalization, Dictionary Skills, and Sentence Sense. The test totals 122 items at the Intermediate I level, 145 items at the Advanced level. The test is not intended to focus on the meaning or understanding of language, but rather on the various more or less mechanical skills associated with correct grammatical and idiomatic use of the English language.

Part A: Usage

The items of Part A: Usage offer two alternative words or phrases as part of a sentence. The test-taker is to indicate, at the Intermediate I level, whether the first choice, the second choice, or neither choice will make the sentence grammatically correct. At the Advanced level a fourth alternative is added, i.e., that both choices are correct. The Detailed Tables for this sub-test are Tables Int-E and Adv-C on pages 24 and 35, for the Intermediate I and Advanced levels, respectively. The results of the analysis by content categories are presented in Table 10.

The only consistency across the two levels seems to be that pronouns present a challenge to the hearing impaired group about equal to that of the standardization group. On both levels verb tenses were the most difficult compared to the standardization group, but the levels of the difference were widely disparate.

Part B: Punctuation

This sub-test is in the format of several paragraphs in which no punctuation at all is included. The Intermediate I passage is a letter; the Advanced passage is a story. Various points in the text are indicated for which the test-taker is to decide the appropriate punctuation. Both levels have three response choices to each item. At the Intermediate I level the first two choices are for specific punctuation as listed, the third choice always being "no punctuation required." At the Advanced level the three options vary at random, including a "no punctuation needed" category and an "optional punctuation" category. The positions of these latter two options vary at random, by contrast with the Intermediate I, where the "no punctuation required" category is always the last option listed. Detailed Tables Int-F and Adv-D on pages 25 and 36 present the basic data for the Intermediate I and Advanced levels, respectively. Some moderate differences occur among the different types of punctuation required, but since the figures are often based on very small numbers of items, they cannot be assumed to be very reliable. Table 11 presents these results.

Part C: Capitalization

This test, with 36 items at the Intermediate I level and 45 at the Advanced level, is in the form of several paragraphs with no punctuation or capitalization included. The Intermediate I passage is an essay; the Advanced passage is a letter. For each underlined word in the text the test-taker is to indicate whether a capital or a small letter is appropriate. Table 12 presents the content analysis for this part of the Language sub-test. The detailed data are presented in Tables Int-G and Adv-E on pages 26 and 37, for the Intermediate I and Advanced levels, respectively. From examination of Table 12 it is clear that there were only small differences between the two types of items and that the difficulty pattern for the two groups of students was very similar.

TABLE 10: CONTENT ANALYSIS OF THE LANGUAGE SUB-TEST, PART A: USAGE, INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
Category	Number of Items	Average Difference	Number of Items	Average Difference
Irregular Verb Forms	9	−11.9	6	− 2.7
Verb Tenses	6	−27.8	5	− 8.8
Verb Agreement with Subject	4	− 8.3	3	+10.0
Other Verb Items	4	−18.0	6	− 4.0
Pronoun Items	7	+ 1.6	8	+ 2.0
Miscellaneous	8	−15.4	10	− 3.1

TABLE 11: CONTENT ANALYSIS OF THE LANGUAGE SUB-TEST, PART B: PUNCTUATION, INTERMEDIATE I AND ADVANCED BATTERIES.

Punctuation	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Comma	6	+ 5.8	9	−3.8
Colon	1	−11.0	–	–
Quotation Marks	3	− 2.0	6	+9.8
Period, Question Mark	3	− 8.3	1	+8.0
No Punctuation Needed	7	− 8.0	2	−5.5
Optional Punctuation	–	–	2	+7.5

TABLE 12: CONTENT ANALYSIS OF THE LANGUAGE SUB-TEST, PART C: CAPITALIZATION, INTERMEDIATE I AND ADVANCED BATTERIES.

Category	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Capital Letter Needed	20	+1.2	26	−0.3
Small Letter Needed	16	−3.1	19	−0.5

Part D: Dictionary Skills

This test, consisting of 10 items at the Intermediate I level and of 24 items at the Advanced level, is intended to assess skill at using a dictionary. The items of the Intermediate I level and the first section of the test at the Advanced level consist of excerpts from a dictionary, showing pronunciation, definitions, etc. Following this is a series of items requiring the use, interpretation and understanding of the dictionary excerpt. The second part of the test at the Advanced level concentrates on the use of the dictionary pronunciation guide to assist in pronunciation of various words. All items at both levels are of the multiple choice type with four response options per item. The results of the analysis by content are presented in Table 13. Detailed Tables Int-H and Adv-F on pages 27 and 38 present the basic data for the Intermediate I and Advanced levels, respectively.

The relative difficulty with pronunciation and sound matching items is predictable and is to be expected in a hearing impaired group. The other figures are less easily explainable and are based on limited numbers of items.

TABLE 13: CONTENT ANALYSIS OF THE LANGUAGE SUB-TEST, PART D: DICTIONARY SKILLS, INTERMEDIATE I AND ADVANCED BATTERIES.

Category	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Choice of Meaning	4	−19.8	9	− 9.8
Pronunciation/Matching Sounds	2	−26.0	8	−13.9
Guide Words	1	+ 2.0	2	+15.5
Other Items	3	−19.0	5	− 5.2

Part E: Sentence Sense

This test consists of strings of words which form one complete sentence, two complete sentences, or an incomplete sentence fragment. The test-taker is to indicate in which category each string of words belongs. The test consists of 18 items at both the Intermediate I and the Advanced levels. The results of the content analysis for this sub-test are presented in Table 14. The detailed data are presented in Tables Int-I and Adv-G on pages 27 and 38, for the Intermediate I and Advanced levels, respectively.

Hearing impaired students clearly did best on items which contained two complete sentences. Since these items also contained the greatest average number of words, it is uncertain whether the length of the item or its content (two sentences) was the attribute being responded to by the hearing impaired students. In comparison, the "incomplete sentence" items were substantially more difficult for the hearing impaired students to recognize correctly. Examination of the data suggests that the hearing impaired students performed very differently on items containing dependent clauses compared to those items which did not include dependent clauses. The results of this examination are shown in Table 15.

It is clear that this task is substantially more difficult for hearing impaired students when the string of words to be identified contains a dependent clause. It is presumably these differences and those reported in Table 14 which contributed to the low correlations between the hearing impaired and the standardization group item difficulty figures for this sub-test (.27 and .30 for the two levels, respectively).

Arithmetic Computation

This test, consisting of 39 items at the Intermediate I level and 41 items at the Advanced level, measures proficiency in the basic computational skills. The items are in multiple choice format, with five response options. Of these, four are specific alternative answers, and the fifth choice is "not given," meaning that the correct answer is not among the first four choices. Since this is an area in which the hearing impaired group generally performed better than the standardization group, it must be recalled that the hearing impaired group who took the Intermediate I battery are, on the average, 5.5 years older than the standardization group for that level. At the Advanced level, the hearing impaired

TABLE 14: CONTENT ANALYSIS OF THE LANGUAGE SUB-TEST, PART E: SENTENCE SENSE, INTERMEDIATE I AND ADVANCED BATTERIES.

Category	Intermediate I			Advanced		
	Number of Items	Average Number Words	Average Difference	Number of Items	Average Number Words	Average Difference
Incomplete Sentence	8	10.0	−20.0	7	11.0	−14.6
One Sentence	5	14.2	−11.4	4	13.3	− 7.3
Two Sentences	5	18.4	+ 7.2	7	14.1	+ 3.4

TABLE 15: FORMAT ANALYSIS OF THE LANGUAGE SUB-TEST, PART E: SENTENCE SENSE, INTERMEDIATE I AND ADVANCED BATTERIES.

Item Type	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Containing Dependent Clauses	4	−19.8	6	−20.7
Without Dependent Clauses	14	− 7.3	12	+ 1.4

students are, on the average, 4.0 years older than the students in the standardization group. Table 16 gives the results of the content analysis of the items in this sub-test. The detailed data are presented in Tables Int-J and Adv-H on pages 28 and 39, for the Intermediate I and Advanced levels, respectively.

Arithmetic Concepts

This sub-test measures understanding of fractions, ratios, arithmetic averages, the meaning of percent, and other concepts related to the broad use of arithmetic principles. Many of the items can be characterized as focusing on use and understanding of mathematical language. All items are in multiple choice format with four response options; the test contains 32 items at the Intermediate I level and 40 items at the Advanced level. Analysis of the content of this test is presented in Table 17. Detailed data appear in Tables Int-K and Adv-I on pages 29 and 40, for the Intermediate I and Advanced levels, respectively.

While the fraction concepts items were particularly difficult for the hearing impaired students at both levels, the test as a whole was more difficult for hearing impaired students at the Intermediate I level than it was at the Advanced level.

If the "language load" of these items is measured simply by counting the number of words in the item stem, and this figure is related to the item difficulty figures and the item difference figures, the results in Table 18 are obtained.

These figures suggest that, while the length of the item is not a significant influence on success for students taking the Advanced battery, it is a significant factor for students at the Intermediate I level. As

TABLE 16: CONTENT ANALYSIS OF THE ARITHMETIC COMPUTATION SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
Category	Number of Items	Average Difference	Number of Items	Average Difference
Addition	7	+ 8.6	5	+ 7.2
Subtraction	7	+ 6.0	4	+ 0.5
Multiplication	9	+14.2	6	+ 1.8
Division	10	+11.2	9	+ 2.2
Fractions	4	+ 4.5	–	–
Money & Decimals	2	+17.5	–	–
Rounding	–	–	3	– 0.7
Percent	–	–	6	–10.7
Equations	–	–	7	+15.3
Exponent	–	–	1	+22.0

TABLE 17: CONTENT ANALYSIS OF THE ARITHMETIC CONCEPTS SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

	Intermediate I		Advanced	
Category	Number of Items	Average Difference	Number of Items	Average Difference
Fraction Concepts	7	–9.6	4	–15.3
Operational Relationships	4	–6.8	12	– 2.2
Place Value	4	–7.0	–	–
Roots and Powers	–	–	3	+20.0
Estimation	–	–	3	+ 1.0
Assorted Content	17	–6.2	18	+ 3.3

TABLE 18: FORMAT ANALYSIS OF THE ARITHMETIC CONCEPTS SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

Correlation of the Number of Words With:	Intermediate I		Advanced	
	Number of Items	Correlation	Number of Items	Correlation
Average Difference	32	+.06	40	−.13
Standardization Item Difficulty	32	−.56*	40	−.17
Hearing Impaired Item Difficulty	32	−.42**	40	−.15

*$p < .01$
**$p < .02$

the length of the item increases, the percentage of students answering the item correctly decreases. This effect, however, is very similar for both groups, and therefore the length of the item does not show a significant relationship to the item difference figures.

Arithmetic Applications

This sub-test measures arithmetic reasoning with problems drawn from life experience – the familiar "word problems." The general reading vocabulary has been kept much below the problem-solving level being measured, and computation difficulty has been controlled so that it is only a minor factor. The test consists of 33 items at the Intermediate I level and of 36 items at the Advanced level. All items are in multiple choice format with four specific answer choices and a fifth option of "not given," meaning that the correct answer is not among the first four options listed. An analysis of the content categories of this sub-test is presented in Table 19. Detailed data appear in Tables Int-L and Adv-J on pages 30 and 41, for the Intermediate I and Advanced levels, respectively.

Examination of the influence of the length of the item, similar to the analysis of the Arithmetic Concepts Sub-test, was undertaken. The results are presented in Table 20. As in the Arithmetic Concepts

TABLE 19: CONTENT ANALYSIS OF THE ARITHMETIC APPLICATIONS SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

Category	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Addition	2	+ 7.0	–	–
Subtraction	6	−13.0	–	–
Multiplication/Division	5	−15.6	–	–
Fractions	3	− 2.3	–	–
Ratio	4	−12.3	–	–
Graph Reading	4	−10.0	8	− 3.8
Problem Analysis	3	+ 5.3	3	− 4.3
Multiple Step Problems	4	− 8.3	2	− 6.5
Measurement	2	− 9.0	4	− 3.8
Rate	–	–	3	−11.0
Proportion	–	–	3	− 4.0
Table Reading	–	–	3	− 5.3
Geometry	–	–	3	+ 6.7
Business Arithmetic	–	–	4	− 5.0
Other Items	–	–	3	− 4.0

TABLE 20: FORMAT ANALYSIS OF THE ARITHMETIC APPLICATIONS SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

Correlation of the Number of Words With:	Intermediate I		Advanced	
	Number of Items	Correlation	Number of Items	Correlation
Average Difference	33	+.20	36	−.05
Standardization Item Difficulty	33	−.41*	36	−.27
Hearing Impaired Item Difficulty	33	−.26	36	−.29**

*$p < .01$

**$p < .05$

Sub-test, the relationship between the number of words in an item and the item difficulty figure is an inverse one; as the length of the item increases, the percentage of students answering the item correctly decreases.

If the influence of language can be seen by this highly simplistic measure of counting the number of words in an item, it seems likely that more sophisticated analysis of the difficulty or complexity of the language in which the items are written would show an even greater degree of relationship between that measure of "language load" and the performance of students on the given sub-test. It is noteworthy that both here and on the Arithmetic Concepts Sub-test, the "language load" measure produces quite similar results for the two groups. In other words, the "language load" of an item seems to influence the performance of the standardization group in approximately the same degree as it influences the performance of the hearing impaired students.

Social Studies

This sub-test consists of two parts. The first, Content, is intended to measure subject matter including geography, history, economics, etc. The second part, Study Skills, involves ability to interpret graphs and tables, maps, and other reference materials. The Intermediate I test includes 24 content items and 25 study skills questions, for a total of 49 items. The Advanced level test is substantially longer, with 52 content items and 40 study skills questions, for a total of 92 items. The content areas of the two levels will be analyzed together in Table 21. Since the study skills materials are very different at the two levels, the data from this part of the Intermediate I level will be analyzed in Table 22, with the corresponding analysis for the Advanced level in Table 23. Detailed data for the Intermediate I level appear in Table Int-M on page 31. Similar data for the Advanced level are presented in Tables Adv-K and Adv-L on pages 42 and 43.

TABLE 21: CONTENT ANALYSIS OF THE SOCIAL STUDIES SUB-TEST, PART A: CONTENT, INTERMEDIATE I AND ADVANCED BATTERIES.

Category	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Geography	7	− 6.6	10	− 3.7
History	5	−14.0	10	− 6.3
Economics	3	− 6.3	7	− 8.1
Industry	3	− 8.7	–	–
Sociology	–	–	9	−10.6
Civics	–	–	14	− 5.2
Other Items	6	− 5.7	2	− 9.0

TABLE 22: CONTENT ANALYSIS OF THE SOCIAL STUDIES SUB-TEST, PART B: STUDY SKILLS, INTERMEDIATE I BATTERY.

Category	Number of Items	Average Difference
Reading a Bar Graph	5	+ 3.6
Reading a Pictograph	5	−11.6
Using a Globe	5	+ 9.2
Reading Map and Legend	10	+10.0

TABLE 23: CONTENT ANALYSIS OF THE SOCIAL STUDIES SUB-TEST, PART B: STUDY SKILLS, ADVANCED BATTERY.

Category	Number of Items	Average Difference
Bar or Line Graph	6	−11.2
References	6	− 8.7
Bibliography	6	+ 2.0
Index Card	6	−10.5
Globe	11	+ 7.6
Political Poster	5	− 2.4

The data shown in these tables sometimes reflect substantial differences which are not readily interpretable (e.g., the pictograph items from Table 22, which are 15 to 22 points below all the other items of Part B, Intermediate I level). It may be that these variations reflect primarily the varying curriculum emphases among these sub-areas of social studies.

Science

The primary objectives of the Science Sub-test are (1) the ability to see the application of the principles of science in our environment and everyday activities, (2) knowledge of the facts and generalizations from the various branches of the natural sciences, and (3) some knowledge of the scientific method. The test consists of 56 items at the Intermediate I level and 60 items at the Advanced level. All items are multiple choice with four response options. The analysis of the content categories of the Science Sub-test is presented in Table 24. Detailed Tables Int-N and Adv-M appear on pages 32 and 44, for the Intermediate I and Advanced levels, respectively.

The differences between the hearing impaired and the standardization group are quite small at the Advanced level, with the sole exception of the electricity and magnetism items. At the Intermediate I level, however, there are many substantial differences, always in favor of the standardization group.

SOME SPECIAL PROBLEMS

Two special problems were encountered during the course of the various analyses already presented. One concerns the positioning of the correct answer in the list of response choices; the other concerns "negative" response options.

The Position of the Correct Response

In the construction of multiple choice tests of the type used in the Stanford Achievement Series,

TABLE 24: CONTENT ANALYSIS OF THE SCIENCE SUB-TEST, INTERMEDIATE I AND ADVANCED BATTERIES.

Category	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
Air & Weather	5	− 9.5	3	+1.0
Astronomy	4	−10.3	4	0.0
Energy & Machines	3	− 3.3	4	+2.0
Earth Science	7	− 5.3	9	−0.8
Animals	8	− 8.6	5	−0.8
Plants	6	− 1.2	7	+2.0
Body	5	−15.8	7	−2.4
Science in Everyday Living	6	−11.5	–	–
Scientific Method	5	−18.8	–	–
Electricity & Magnetism	–	–	5	−8.4
Other Items	7	−11.0	16	−4.1

one of the technical considerations is the placement of the correct answer among the listed alternatives. In general, the attempt is made to randomize the position of the correct answer, so that in a given test the correct answer will appear in the various positions about an equal number of times. Thus, in a 60-item test with four alternatives for each item, the correct answer will be in the first position about 15 times, in the second position about 15 times, etc. It is to be expected, then, that the performance of a group of students would be about the same on the group of items with correct answers in the first position, as on the group of items with correct answers in the second position, etc. This, however, was not the case with the hearing impaired students in this study.

Using the difference scores employed throughout this study, it was observed that in most instances substantial differences occurred in the hearing impaired students' performance, relative to the standardization group's performance, according to the position in which the correct response appeared. For example, on the Intermediate I Paragraph Meaning Sub-test, the difference score was – 20.0 for items with the correct answer in the second position and – 5.6 for items with the correct answer in the fourth position. Table 25 presents the difference scores for each sub-test of both batteries according to the position of the correct answer. The final (rightmost) column, labeled "greatest difference," gives the largest difference in score observed between two of the response positions for that sub-test. In other words, all pairs of difference scores were compared, and the greatest observed difference among these pairs was reported.

While the specific position showing the best, or the poorest, performance among hearing impaired students varied from sub-test to sub-test, the overall results presented in Table 26 continue to reflect differences related to positioning of the correct answer. The entries in Table 26 indicate the number of sub-tests (or sub-test parts such as punctuation) in which each response position showed best and poorest performance for hearing impaired students relative to the standardization group.

For the Intermediate I battery, then, the first two positions produced the best performance 12 times out of 14. In the Advanced battery, positions 1 and 4 were best 11 times out of 13. Poorest performance was more evenly distributed in both cases.

These results are not readily interpretable with any confidence. A number of possibilities come to mind immediately. Students may be unaware of the need to consider all alternatives before responding, an informational lack. They may be unwilling to consider all the alternatives, an "impulsive" response style, a behavioral lack. Some experimental study and the evaluation of various kinds of test-taking training will be needed to clarify the issues raised by these findings.

"Negative" Responses

The second special problem noted during the conduct of the foregoing analyses is that of the hearing impaired students' response to items in which the correct answer is a "negative" answer. Such "negative" response options occur in the following sub-tests:

Language, Part A: Usage
Language, Part B: Punctuation
Language, Part E: Sentence Sense
Arithmetic Computation
Arithmetic Applications.

The exact form of the "negative" option varies from sub-test to sub-test, including the following:

"N" (neither alternative is correct)
"NP" (no punctuation is needed)
"NG" (the correct answer is not given).

Hearing impaired students tend to do substantially less well than the standardization group on items where the correct answer is "negative" as described above. The results of analysis of this phenomenon are presented in Table 27.

Explanations for this occurrence must again be speculative, but may involve cognitive style differences, perhaps related to lack of practice with similar "negative" alternatives during ordinary classroom drill. Again, full explication of the roots of this phenomenon must await experimental studies focused on this issue.

SUMMARY

This study has investigated differential patterns of the performance of hearing impaired students compared with the hearing standardization group on whom the Stanford Achievement Test was normed. The basic statistic used was the "difference" figure, the arithmetic difference between the item difficulty indexes for the two student groups. In this way the intra-test variations of difficulty are controlled, and the figures indicate only the degree to which the performance of the hearing impaired students differed from that of the standardization group. These differences were investigated for the content sub-

TABLE 25: GROUP DIFFERENCE SCORES ACCORDING TO POSITION OF THE CORRECT ANSWER, INTERMEDIATE I AND ADVANCED BATTERIES.

Sub-test and Level	Position #1		Position #2		Position #3		Position #4		Position #5		Greatest Difference
	No. of Items	Average Difference	No. of Items	Average Difference	No. of Items	Average Difference	No. of Items	Average Difference	No. of Items	Average Difference	
Word Meaning											
Intermediate I	10	−20.6	10	−14.1	9	−14.9	9	−15.8			6.5
Paragraph Meaning											
Intermediate I	15	−12.9	15	−20.0	16	−12.9	14	− 5.6			14.4
Advanced	15	− 2.4	15	− 7.9	16	− 7.3	14	− 7.2			5.5
Spelling											
Intermediate I	13	+ 7.9	11	+15.4	14	+11.4	12	+10.5			7.5
Advanced	14	+12.8	15	+ 7.7	15	+10.1	14	+10.9			5.1
Language Total											
Intermediate I	47	− 4.5	40	− 3.8	34	−15.0	1	−24.0			20.2
Advanced	45	− 2.8	49	− 0.5	25	− 7.3	16	− 1.5			6.8
A: Usage											
Intermediate I	12	−15.2	11	− 6.4	15*	−15.9*					9.5
Advanced	12	− 4.0	11	− 0.3	4	− 8.0	11*	+ 0.8*			8.8
B: Punctuation											
Intermediate I	9	+ 1.7	4	+ 7.0	7*	− 8.0*					15.0
Advanced	7	− 1.3	6	+ 2.7	7	+ 4.3					5.6
C: Capitalization											
Intermediate I	20	+ 1.2	16	− 3.1							4.3
Advanced	26	− 0.3	19	− 0.5							0.2
D: Dictionary Skills											
Intermediate I	1	−12.0	4	−23.8	4	−13.8	1	−24.0			12.0
Advanced	6	− 5.0	6	− 8.8	7	−11.1	5	− 6.6			6.1
E: Sentence Sense											
Intermediate I	5	−11.4	5	+ 7.2	8*	−20.0*					27.2
Advanced	4	− 7.3	7	+ 3.4	7*	−14.6*					18.0
Arithmetic Computation											
Intermediate I	9	+14.6	9	+11.6	9	+ 6.3	8	+12.4	4*	+ 1.0*	13.6
Advanced	10	+ 2.3	9	+ 3.0	8	− 0.8	10	+ 7.7	4*	+ 2.3*	8.5
Arithmetic Concepts											
Intermediate I	9	− 4.8	8	− 4.3	8	−10.6	7	− 9.3			6.3
Advanced	10	− 5.9	11	+ 2.4	9	+ 1.8	10	+ 5.3			11.2
Arithmetic Applications											
Intermediate I	6	− 9.3	7	−12.4	8	− 8.6	7	+ 3.0	5*	−20.0*	23.0
Advanced	8	−10.9	8	− 5.0	8	− 2.9	7	+ 3.6	5*	− 4.6*	14.5
Social Studies Total											
Intermediate I	12	+ 6.2	13	− 5.8	12	− 2.2	12	− 5.2			14.0
Advanced	22	− 3.5	23	− 7.5	23	− 6.4	24	− 1.9			5.6
A: Content											
Intermediate I	6	+ 0.2	6	−12.8	6	−10.0	6	− 9.8			13.0
Advanced	13	− 4.4	13	− 9.2	13	− 8.5	13	− 4.2			5.0
B: Study Skills											
Intermediate I	6	+12.2	7	+ 0.3	6	+ 5.7	6	− 0.5			12.7
Advanced	9	− 2.1	10	− 5.2	10	− 3.7	11	+ 0.8			6.0
Science											
Intermediate I	14	− 5.6	14	− 6.6	13	−10.5	15	−11.1			4.5
Advanced	14	− 4.7	17	− 2.8	14	− 1.4	15	+ 1.5			6.2

*This response choice is a "negative" option (see text for explanation).

TABLE 26: FREQUENCY OF BEST AND POOREST RELATIVE PERFORMANCE FOR HEARING IMPAIRED STUDENTS, BY SUB-TEST AND PARTS OF SUB-TESTS, ACCORDING TO POSITION OF CORRECT RESPONSE, INTERMEDIATE I AND ADVANCED BATTERIES.

Position of Correct Response	Intermediate I		Advanced	
	Best	Poorest	Best	Poorest
1	5	2	4	4
2	7	3	1	5
3	0	4	1	4
4	2	3	7	0
5	0	2	0	0

TABLE 27: RELATIVE PERFORMANCE OF HEARING IMPAIRED STUDENTS ON "NEGATIVE" AND NON-"NEGATIVE" RESPONSES, INTERMEDIATE I AND ADVANCED BATTERIES.

Item Type Correct Answer	Intermediate I		Advanced	
	Number of Items	Average Difference	Number of Items	Average Difference
"Negative"	39	−14.1	29	−4.1
Non-"Negative" (All Other Items)	440	− 4.6	503	−1.1

categories of each of the sub-tests of the Intermediate I and Advanced level batteries. The format of the study emphasizes presentations of data reflecting the differences in performance which occurred; interpretations of these differences are minimal in this text, since such interpretations require experimental study for confirmation or disconfirmation. The value of the data presented here lies in their ability to generate hypotheses to explain the observed differences and to stimulate research to test those hypotheses. The data can also be of value for pointing to specific types of academic content in which, for whatever reason, hearing impaired students performed more or less adequately than their hearing peers. The potential value of such information for curriculum planning and classroom activity is obvious.

For the reader who wants to investigate in greater detail the performance, or relative performance, of hearing impaired students on specific items, the complete basic data on which this study was based are included in Detailed Tables Int-A through Adv-M. For each item of each sub-test of the two batteries, three figures are given: the item difficulty index for the hearing impaired group; the item difficulty index for the standardization group; and the arithmetic difference between the two figures. In addition, the content category for each item is listed.

TABLE Int-A: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, WORD MEANING SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Part of Speech of Correct Choice	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Adverb	62	92	−30
2	Verb	70	86	−16
3	Noun	66	91	−25
4	Noun	98	96	+ 2
5	Noun	78	70	+ 8
6	Noun	65	86	−21
7	Adjective	40	76	−36
8	Verb	40	73	−33
9	Verb	62	88	−26
10	Adjective	34	67	−33
11	Adjective	55	63	− 8
12	Adjective	30	70	−40
13	Noun	51	67	−16
14	Adjective	34	75	−41
15	Noun	53	71	−18
16	Adjective	42	71	−29
17	Noun	58	72	−14
18	Adjective	19	61	−42
19	Noun	31	65	−34
20	Adjective	29	48	−19
21	Verb	27	54	−27
22	Verb	32	50	−18
23	Noun	27	57	−30
24	Verb	25	37	−12
25	Noun	43	36	+ 7
26	Adjective	46	44	+ 2
27	Verb	25	43	−18
28	Adjective	20	37	−17
29	Verb	22	28	− 6
30	Adjective	30	34	− 4
31	Adjective	21	38	−17
32	Noun	18	30	−12
33	Verb	14	21	− 7
34	Verb	13	13	0
35	Adjective	13	16	− 3
36	Verb	25	18	+ 7
37	Adjective	20	17	+ 3
38	Verb	15	15	0

TABLE Int-B: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, PARAGRAPH MEANING SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Classification	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Comprehension – literal	61	84	–23
2	Comprehension – inferential	46	88	–42
3	Comprehension – inferential	16	82	–66
4	Comprehension – literal	44	67	–23
5	Comprehension – inferential	13	70	–57
6	Comprehension – literal	51	83	–32
7	Comprehension – literal	33	66	–33
8	Comprehension – literal	80	90	–10
9	Comprehension – literal	62	90	–28
10	Inference	66	87	–21
11	Comprehension – inferential	47	82	–35
12	Comprehension – inferential	36	71	–35
13	Comprehension – literal	32	73	–41
14	Comprehension – literal	44	70	–26
15	Comprehension – literal	57	79	–22
16	Comprehension – literal	62	61	+ 1
17	Comprehension – literal	49	55	– 6
18	Comprehension – inferential	35	65	–30
19	Comprehension – inferential	81	83	– 2
20	Comprehension – literal	40	66	–26
21	Comprehension – inferential	26	59	–33
22	Comprehension – inferential	54	69	–15
23	Comprehension – inferential	23	55	–32
24	Inference	44	67	–23
25	Inference	53	68	–15
26	Inference	23	54	–31
27	Inference	31	71	–40
28	Comprehension – inferential	75	65	+10
29	Inference	44	57	–13
30	Comprehension – inferential	29	27	+ 2
31	Comprehension – literal	66	60	+ 6
32	Comprehension – literal	78	68	+10
33	Comprehension – literal	49	59	–10
34	Comprehension – inferential	68	80	–12
35	Comprehension – literal	42	43	– 1
36	Inference	38	39	– 1
37	Comprehension – literal	32	41	– 9
38	Inference	25	38	–13
39	Comprehension – inferential	54	57	– 3
40	Inference	30	31	– 1
41	Inference	35	38	– 3
42	Comprehension – inferential	43	54	–11
43	Comprehension – inferential	10	16	– 6
44	Comprehension – inferential	38	37	+ 1
45	Comprehension – inferential	21	30	– 9
46	Comprehension – inferential	16	24	– 8
47	Comprehension – inferential	43	39	+ 4
48	Comprehension – inferential	41	33	+ 8
49	Comprehension – inferential	25	18	+ 7
50	Inference	25	30	– 5
51	Comprehension – literal	51	53	– 2
52	Comprehension – inferential	29	32	– 3
53	Comprehension – literal	47	35	+12
54	Comprehension – literal	36	27	+ 9
55	Comprehension – inferential	12	16	– 4
56	Inference	28	27	+ 1
57	Comprehension – inferential	18	18	0
58	Comprehension – inferential	13	13	0
59	Comprehension – inferential	15	15	0
60	Comprehension – inferential	30	20	+10

TABLE Int-C: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SPELLING SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Misspelled Word and Part of Speech Classification [1]		Item Difficulty Values		Difference
			Hearing Impaired Students	Hearing Students	
1	gatherd	(verb)	86	83	+ 3
2	dremed	(verb)	85	86	− 1
3	packege	(noun)	81	81	0
4	worng	(adjective)	91	83	+ 8
5	lafing	(verb)	88	85	+ 3
6	bord	(noun)	89	81	+ 8
7	ascape	(verb)	89	78	+11
8	flud	(noun)	91	83	+ 8
9	slowely	(adverb)	89	80	+ 9
10	wate	(verb)	87	79	+ 8
11	adress	(noun)	92	81	+11
12	kichen	(noun)	92	83	+ 9
13	slideing	(verb)	79	81	− 2
14	allthough	(conjunction)	80	79	+ 1
15	replie	(verb)	85	76	+ 9
16	screemed	(verb)	71	58	+13
17	westren	(adjective)	84	74	+10
18	doller	(noun)	89	73	+16
19	everbody	(pronoun)	84	75	+ 9
20	hoby	(noun)	84	69	+15
21	minuts	(noun)	85	75	+10
22	peice	(noun)	87	76	+11
23	dich	(noun)	91	66	+25
24	suposed	(verb)	85	71	+14
25	developped	(verb)	48	38	+10
26	inocent	(adjective)	77	64	+13
27	colection	(noun)	57	42	+15
28	heavey	(adjective)	84	68	+16
29	autum	(noun)	86	76	+10
30	truely	(adverb)	70	56	+14
31	offerred	(verb)	67	52	+15
32	liebility	(noun)	61	62	− 1
33	preasent	(noun)	74	69	+ 5
34	avalable	(adjective)	50	47	+ 3
35	cryed	(verb)	83	61	+22
36	visting	(verb)	57	50	+ 7
37	slippry	(adjective)	80	54	+26
38	realy	(adverb)	78	61	+17
39	begining	(verb)	51	46	+ 5
40	possable	(adjective)	49	40	+ 9
41	dimond	(noun)	72	41	+31
42	sholders	(noun)	57	33	+24
43	listend	(verb)	56	47	+ 9
44	sevral	(adjective)	68	53	+15
45	appeard	(verb)	53	40	+13
46	continuelly	(adverb)	31	40	− 9
47	favorate	(adjective)	31	12	+19
48	desided	(verb)	53	31	+22
49	foriegn	(adjective)	46	23	+23
50	carring	(verb)	45	30	+15

[1] Any words that could be more than one part of speech were classified by the most common usage.

DATA FOR TABLE Int.-D, WORD STUDY SKILLS SUB-TEST, INTERMEDIATE I BATTERY, FORM W, ARE NOT REPORTED DUE TO ITS UNSUITABILITY FOR HEARING IMPAIRED STUDENTS.

TABLE Int-E: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART A: USAGE, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Irregular verb forms	67	93	−26
2	Verbs: Correct form of subjunctive mood	59	91	−32
3	Verb tenses	64	89	−25
4	Substandard corruption form	76	86	−10
5	Double negative	67	82	−15
6	Pronouns: Substandard corruption form	55	73	−18
7	Correct form of pronoun	72	85	−13
8	Irregular verb forms	59	86	−27
9	Irregular verb forms	69	79	−10
10	Verb agreement with subject	63	87	−24
11	Verbs: Substandard corruption form	74	86	−12
12	Verbs: Correct form of subjunctive mood	63	69	− 6
13	Possessive pronoun	91	81	+10
14	Spelling	43	87	−44
15	Word choice involving verbs	56	79	−23
16	Verb tenses	24	68	−44
17	Verb agreement with subject	30	66	−36
18	Irregular verb forms	27	67	−40
19	Verb tenses	48	79	−31
20	Verbs: Substandard corruption form	36	72	−36
21	Verb agreement with subject	67	52	+15
22	Pronoun nominative case	41	49	− 8
23	Verb tenses	33	55	−22
24	Word choice involving verbs	37	48	−11
25	Spelling	88	61	+27
26	Verb agreement with subject	56	44	+12
27	Pronoun nominative case	47	55	− 8
28	Word choice involving verbs	10	35	−25
29	Irregular verb forms	27	34	− 7
30	Pronoun objective case	46	30	+16
31	Irregular verb forms	37	40	− 3
32	Verbs: Double negative	25	40	−15
33	Irregular verb forms	27	28	− 1
34	Irregular verb forms	32	19	+13
35	Irregular verb forms	22	28	− 6
36	Pronoun nominative case	31	26	+ 5
37	Verb tenses	16	36	−20
38	Correct form of pronoun	29	20	+ 9

TABLE Int-F: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART B: PUNCTUATION, INTERMEDIATE I BATTERY, FORM W.

Item Number	Rule of Punctuation	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
39	None required: In street address	77	70	+ 7
40	Comma: Between city and state	92	87	+ 5
41	None required: After date	55	61	− 6
42	None required	51	56	− 5
43	None required: After state in address	46	44	+ 2
44	Colon: After business salutation	29	40	−11
45	None required: In middle of sentence	78	73	+ 5
46	Comma: Separating clauses	44	28	+16
47	Period: End of sentence	45	40	+ 5
48	Comma: After conjunctive adverb	46	39	+ 7
49	Comma: Separating clauses	38	36	+ 2
50	None required: In middle of sentence	13	61	−48
51	Period: End of sentence	90	84	+ 6
52	None required: In middle of sentence	40	51	−11
53	Comma: After complimentary close	88	84	+ 4
54	Comma: After direct address	35	34	+ 1
55	Quotation marks: Close quotation	53	50	+ 3
56	Period: End of sentence	60	46	+14
57	Quotation marks: Close quotation	26	35	− 9
58	Quotation marks: Close quotation	57	57	0

TABLE Int-G: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART C: CAPITALIZATION, INTERMEDIATE I BATTERY, FORM W.

Item Number	Capitalization Rule	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
59	No caps: Common noun, middle of sentence	26	51	−25
60	Name of a country	94	96	− 2
61	Name of a city	77	80	− 3
62	No caps: Common noun, middle of sentence	79	84	− 5
63	Epithet for particular flag (in quotes)	89	86	+ 3
64	No caps: Common noun, middle of sentence	82	89	− 7
65	No caps: Common noun, middle of sentence	71	61	+10
66	Epithet for particular flag (in quotes)	94	90	+ 4
67	Epithet for particular flag (in quotes)	92	90	+ 2
68	Epithet for particular flag (in quotes)	88	85	+ 3
69	Epithet for particular flag (in quotes)	83	85	− 2
70	Epithet for particular flag (in quotes)	73	67	+ 6
71	No caps: Conjunction in middle of title	89	85	+ 4
72	No caps: Geographical direction	69	71	− 2
73	Proper name of river	91	90	+ 1
74	Proper name of river	60	58	+ 2
75	Name of a city	89	91	− 2
76	No caps: Common noun, middle of sentence	93	90	+ 3
77	No caps: Adjective, middle of sentence	53	68	−15
78	No caps: Common noun, middle of sentence	64	78	−14
79	No caps: Common noun, middle of sentence	84	86	− 2
80	Name of a country	90	92	− 2
81	No caps: Common noun, middle of sentence	50	48	+ 2
82	Title of musical work (in quotes)	92	89	+ 3
83	No caps: Article in middle of title	88	84	+ 4
84	Nationality	90	91	− 1
85	Nationality	88	89	− 1
86	Title of musical work (in quotes)	86	77	+ 9
87	No caps: Common noun, middle of sentence	88	82	+ 6
88	Title of specific governing body	54	56	− 2
89	No caps: Common noun, middle of sentence	66	59	+ 7
90	Name of a language	89	85	+ 4
91	Name of a language	88	85	+ 3
92	No caps: Common noun, middle of sentence	70	78	− 8
93	Proper name; body of water	81	82	− 1
94	No caps: Common noun, end of sentence	78	85	− 7

TABLE Int-H: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART D: DICTIONARY SKILLS, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
95	Choice of meaning	38	83	−45
96	Choice of meaning	41	43	− 2
97	Matching sound	49	85	−36
98	Interpreting a definition	30	36	− 6
99	Interpreting a definition	15	27	−12
100	Choice of meaning	36	44	− 8
101	Choice of meaning	28	52	−24
102	Interpreting a definition	39	78	−39
103	Guide words	46	44	+ 2
104	Matching sound	24	40	−16

TABLE Int-I: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART E: SENTENCE SENSE, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
105	Incomplete sentence	51	34	+17
106	Incomplete sentence	49	80	−31
107	Two complete sentences	61	55	+ 6
108	One complete sentence	73	81	− 8
109	Incomplete sentence	46	82	−36
110	Two complete sentences	62	49	+13
111	One complete sentence	32	66	−34
112	Incomplete sentence	15	58	−43
113	Incomplete sentence	73	81	− 8
114	Incomplete sentence	18	55	−37
115	Two complete sentences	59	45	+14
116	One complete sentence	60	74	−14
117	Two complete sentences	47	48	− 1
118	One complete sentence	59	61	− 2
119	Incomplete sentence	29	44	−15
120	Two complete sentences	59	55	+ 4
121	Incomplete sentence	50	57	− 7
122	One complete sentence	59	58	+ 1

TABLE Int-J: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, ARITHMETIC COMPUTATION SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Addition: 2d + 2d[1]	94	88	+ 6
2	Subtraction: 3d – 2d*	84	87	– 3
3	Multiplication: 2d x 1d	95	94	+ 1
4	Addition: 3d + 3d*	89	84	+ 5
5	Multiplication: 2d x 1d	94	91	+ 3
6	Addition: Open sentence (2d + 2d)*	89	86	+ 3
7	Subtraction: 3d – 3d*	78	73	+ 5
8	Addition: Open Sentence (2d + 2d + 1d)*	89	82	+ 7
9	Subtraction: 3d – 3d*	80	67	+13
10	Multiplication: 3d x 1d*	85	75	+10
11	Subtraction: 3d – 3d*	82	75	+ 7
12	Division: 3d ÷ 1d (zero in dividend)	81	84	– 3
13	Addition: 2d + 3d + 2d + 3d*	74	59	+15
14	Multiplication: Open sentence (3d x 1d)*	87	78	+ 9
15	Addition: 3d + 3d + 3d*	80	71	+ 9
16	Addition: 4d + 4d + 4d*	72	57	+15
17	Money: Subtraction (4d – 4d)*	73	58	+15
18	Division: Open sentence (3d ÷ 1d)*	82	80	+ 2
19	Subtraction: 4d – 3d*	64	61	+ 3
20	Subtraction: 5d – 4d*	61	52	+ 9
21	Division: 3d ÷ 1d*	76	68	+ 8
22	Fraction*	65	63	+ 2
23	Fraction	64	68	– 4
24	Multiplication: 3d x 1d*	62	54	+ 8
25	Subtraction: 5d – 4d*	62	54	+ 8
26	Fraction*	57	54	+ 3
27	Division: 3d ÷ 1d*	63	48	+15
28	Division: 4d ÷ 1d*	61	47	+14
29	Division: 4d ÷ 1d*	61	49	+12
30	Division: 4d ÷ 1d; zero in quotient	42	38	+ 4
31	Division: 4d ÷ 1d; zero in quotient	45	40	+ 5
32	Multiplication: 3d x 2d; zero*	62	44	+18
33	Division: 4d ÷ 1d*	59	42	+17
34	Division: 2d ÷ 2d	67	29	+38
35	Multiplication: 3d x 3d; zeroes*	57	27	+30
36	Money: Multiplication (3d x 2d)*	47	27	+20
37	Multiplication: 3d x 2d*	46	23	+23
38	Fraction with numerator larger than 1	36	19	+17
39	Multiplication: 3d x 3d; zero*	45	19	+26

[1] The letter "d" stands for the number of digits in the computation, e.g., 2d + 2d means 2 digits plus 2 digits (32 + 86).

*Items followed by an asterisk (*) involve regrouping (borrowing or carrying).

TABLE Int-K: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, ARITHMETIC CONCEPTS SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Place value	51	74	−23
2	Roman numeral	54	68	−14
3	Subtraction terms	72	74	− 2
4	Fraction concept	64	54	+10
5	Number sentence	76	64	+12
6	Place value	59	63	− 4
7	Multiplication vocabulary	40	61	−21
8	Extending number series	56	67	−11
9	Roman numeral	60	76	−16
10	Place value	64	66	− 2
11	Number sentence	65	58	+ 7
12	Fraction concept	61	60	+ 1
13	Time	78	69	+ 9
14	Fraction concept	56	77	−21
15	Fraction concept	42	63	−21
16	Reading numerals	60	61	− 1
17	Size of number	36	60	−24
18	Extending number series	62	68	− 6
19	Fraction concept	56	56	0
20	Reading numerals	61	62	− 1
21	Fraction concept	33	66	−33
22	Operational relationship in addition	26	49	−23
23	Place value	48	47	+ 1
24	Operational relationship in multiplication	27	41	−14
25	Percent	45	54	− 9
26	Operational relationship in division	35	37	− 2
27	Estimation	53	47	+ 6
28	Fraction concept	31	34	− 3
29	Operational relationship in multiplication	53	41	+12
30	Finding average	26	41	−15
31	Directional number	22	31	− 9
32	Reasoning	26	36	−10

TABLE Int-L: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, ARITHMETIC APPLICATIONS SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Addition (money)	84	83	+ 1
2	Subtraction	72	79	− 7
3	Subtraction	47	71	−24
4	Subtraction	54	61	− 7
5	2-step problem (addition, money)	29	51	−22
6	Division (measurement)	33	47	−14
7	Reading a graph	92	88	+ 4
8	Reading a graph	66	78	−12
9	Reading a graph	35	47	−12
10	Reading a graph	36	56	−20
11	Ratio (measurement)	55	74	−19
12	Measurement	51	69	−18
13	Division (finding average)	42	55	−13
14	Division (money)	38	68	−30
15	Subtraction (money)	56	74	−18
16	Addition (money)	82	69	+13
17	Subtraction (money)	64	59	+ 5
18	Subtraction (money)	40	67	−27
19	Division	49	60	−11
20	Multiplication	45	55	−10
21	Fractions	15	42	−27
22	Fractions	50	32	+18
23	Ratio	35	39	− 4
24	Measurement (multiplication, money)	52	52	0
25	Ratio (money)	34	44	−10
26	Fractions	45	43	+ 2
27	3-step problem (addition, multiplication, money)	37	35	+ 2
28	2-step problem (addition, subtraction, money)	41	38	+ 3
29	2-step problem (finding average)	19	35	−16
30	Ratio (money)	24	40	−16
31	Problem analysis	43	32	+11
32	Problem analysis	26	32	− 6
33	Problem analysis	35	24	+11

TABLE Int-M: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SOCIAL STUDIES SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Geography	89	95	− 6
2	Economics	83	86	− 3
3	Geography	71	74	− 3
4	History	46	75	−29
5	Clothing	80	65	+15
6	Occupation	48	65	−17
7	Industry	56	73	−17
8	Food	57	49	+ 8
9	History	27	39	−12
10	Industry	64	57	+ 7
11	Economics	43	61	−18
12	Industry	33	49	−16
13	Geography	52	53	− 1
14	Sociology and Civics	28	47	−19
15	History	28	54	−26
16	History	50	53	− 3
17	Geography	38	49	−11
18	Communication	50	47	+ 3
19	Geography	40	57	−17
20	Economics	31	29	+ 2
21	Shelter	4	28	−24
22	History	16	16	0
23	Geography	17	29	−12
24	Geography	25	21	+ 4
25	Reading a bar graph	86	90	− 4
26	Reading a bar graph	88	84	+ 4
27	Reading a bar graph	77	73	+ 4
28	Reading a bar graph	54	62	− 8
29	Reading a bar graph	57	35	+22
30	Reading a pictograph	81	76	+ 5
31	Reading a pictograph	53	72	−19
32	Reading a pictograph	49	61	−12
33	Reading a pictograph	18	33	−15
34	Reading a pictograph	21	38	−17
35	Using a globe	87	72	+15
36	Using a globe	82	67	+15
37	Using a globe	67	65	+ 2
38	Using a globe	74	48	+26
39	Using a globe	13	25	−12
40	Reading map and legend	91	79	+12
41	Reading map and legend	86	76	+10
42	Reading map and legend	78	70	+ 8
43	Reading map and legend	71	60	+11
44	Reading map and legend	63	43	+20
45	Reading map and legend	60	43	+17
46	Reading map and legend	37	36	+ 1
47	Reading map and legend	36	28	+ 8
48	Reading map and legend	30	28	+ 2
49	Reading map and legend	37	26	+11

TABLE Int-N: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SCIENCE SUB-TEST, INTERMEDIATE I BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Differences
		Hearing Impaired Students	Hearing Students	
1	Science in everyday living: industry	89	95	− 6
2	Light	86	93	− 7
3	Chemistry	76	90	−14
4	Astronomy	69	77	− 8
5	Astronomy	65	75	−10
6	Science in everyday living: industry	70	91	−21
7	Weather	51	91	−40
8	Animals	36	78	−42
9	Scientific method	69	83	−14
10	Safety	24	72	−48
11	Energy and machines	73	79	− 6
12	Health	47	61	−14
13	Science in everyday living	70	72	− 2
14	Plants	76	72	+ 4
15	Earth science	57	69	−12
16	Animals	56	59	− 3
17	Animals	72	75	− 3
18	Health	68	67	+ 1
19	Air	62	54	+ 8
20	Energy and machines	59	74	−15
21	Magnetism	59	68	− 9
22	Astronomy	61	75	−14
23	Science in everyday living	35	68	−33
24	Conservation	37	58	−21
25	Science in everyday living	62	55	+ 7
26	Air	56	74	−18
27	Animals	65	61	+ 4
28	Earth science	60	54	+ 6
29	Earth science	51	61	−10
30	Famous scientist	42	49	− 7
31	Scientific method	43	79	−36
32	Scientific method	30	49	−19
33	Electricity	39	54	−15
34	Chemistry	39	43	− 4
35	Food, health	39	42	− 3
36	Earth science	20	39	−19
37	Science in everyday living: industry	36	50	−14
38	Animals	26	51	−25
39	Plants	29	40	−11
40	Plants	51	39	+12
41	Animals	31	35	− 4
42	Plants	24	37	−13
43	Earth science	17	34	−17
44	Plants	25	35	−10
45	Energy and machines	70	59	+11
46	Earth science	38	38	0
47	Scientific method	15	33	−18
48	Animals	25	30	− 5
49	Food, health	20	35	−15
50	Astronomy	27	36	− 9
51	Plants	34	23	+11
52	Weather	24	22	+ 2
53	Scientific method	12	19	− 7
54	Animals	28	19	+ 9
55	Earth science	35	20	+15
56	Air and weather	15	14	+ 1

TABLE Adv-A: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, PARAGRAPH MEANING SUB-TEST, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Comprehension – literal	60	84	−24
2	Comprehension – inferential	53	79	−26
3	Comprehension – literal	63	79	−16
4	Comprehension – inferential	73	76	− 3
5	Comprehension – literal	73	78	− 5
6	Inference	43	48	− 5
7	Comprehension – inferential	64	66	− 2
8	Comprehension – inferential	63	75	−12
9	Comprehension – inferential	80	85	− 5
10	Comprehension – literal	41	52	−11
11	Comprehension – literal	69	72	− 3
12	Comprehension – inferential	56	76	−20
13	Comprehension – inferential	57	65	− 8
14	Comprehension – literal	44	56	−12
15	Comprehension – literal	28	56	−28
16	Comprehension – inferential	62	59	+ 3
17	Comprehension – inferential	48	62	−14
18	Comprehension – literal	59	84	−25
19	Inference	41	64	−23
20	Comprehension – inferential	75	71	+ 4
21	Comprehension – literal	47	58	−11
22	Inference	72	79	− 7
23	Comprehension – inferential	56	51	+ 5
24	Comprehension – inferential	76	65	+11
25	Comprehension – inferential	73	70	+ 3
26	Comprehension – inferential	47	71	−24
27	Inference	28	49	−21
28	Inference	45	43	+ 2
29	Comprehension – inferential	62	57	+ 5
30	Comprehension – inferential	47	54	− 7
31	Inference	53	72	−19
32	Comprehension – literal	43	56	−13
33	Comprehension – literal	81	80	+ 1
34	Comprehension – literal	46	59	−13
35	Comprehension – inferential	32	50	−18
36	Comprehension – literal	46	53	− 7
37	Comprehension – literal	40	60	−20
38	Comprehension – inferential	60	66	− 6
39	Comprehension – literal	60	52	+ 8
40	Comprehension – literal	51	45	+ 6
41	Inference	30	30	0
42	Inference	64	68	− 4
43	Comprehension – inferential	77	60	+17
44	Comprehension – inferential	21	22	− 1
45	Comprehension – inferential	41	57	−16
46	Comprehension – inferential	37	47	−10
47	Comprehension – inferential	31	42	−11
48	Comprehension – inferential	24	45	−21
49	Inference	34	39	− 5
50	Inference	58	43	+15
51	Inference	36	19	+17
52	Inference	30	39	− 9
53	Inference	39	34	+ 5
54	Inference	39	38	+ 1
55	Inference	25	27	− 2
56	Comprehension – inferential	14	19	− 5
57	Inference	26	22	+ 4
58	Inference	34	30	+ 4
59	Inference	24	18	+ 6
60	Inference	25	22	+ 3

TABLE Adv-B: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SPELLING SUB-TEST, ADVANCED BATTERY, FORM W.

Item Number	Misspelled Word and Part of Speech Classification[1]		Item Difficulty Values		Difference
			Hearing Impaired Students	Hearing Students	
1	purfume	(noun)	91	85	+ 6
2	soceity	(noun)	79	70	+ 9
3	amoung	(preposition)	94	80	+14
4	glacail	(adjective)	88	89	− 1
5	thourogh	(adjective)	93	79	+14
6	tragidy	(noun)	82	72	+10
7	ambitius	(adjective)	69	69	0
8	companys	(noun)	70	67	+ 3
9	interrim	(noun)	83	81	+ 2
10	perrenial	(adjective)	82	71	+11
11	condem	(verb)	51	53	− 2
12	pasttime	(noun)	72	75	− 3
13	curcumstance	(noun)	67	56	+11
14	esential	(adjective)	73	64	+ 9
15	tennacity	(noun)	60	58	+ 2
16	eficiency	(noun)	62	52	+10
17	carring	(verb)	82	76	+ 6
18	emergancy	(noun)	62	54	+ 8
19	begining	(verb)	73	65	+ 8
20	temperture	(noun)	70	66	+ 4
21	cabinate	(noun)	71	60	+11
22	servent	(noun)	72	55	+17
23	ordinarally	(adverb)	52	47	+ 5
24	divison	(noun)	62	59	+ 3
25	interupt	(verb)	68	59	+ 9
26	schedualed	(verb)	69	46	+23
27	resistence	(noun)	63	55	+ 8
28	comunities	(noun)	76	66	+10
29	recomended	(verb)	51	45	+ 6
30	convienient	(adjective)	63	54	+ 9
31	forword	(adverb)	57	47	+10
32	equiped	(verb)	51	39	+12
33	originaly	(adverb)	71	58	+13
34	favorible	(adjective)	69	43	+26
35	disect	(verb)	61	58	+ 3
36	diciple	(noun)	78	58	+20
37	religous	(adjective)	68	58	+10
38	simpathy	(noun)	68	45	+23
39	discribe	(verb)	64	44	+20
40	parisite	(noun)	63	55	+ 8
41	apperance	(noun)	60	50	+10
42	advertisment	(noun)	38	33	+ 5
43	cancelation	(noun)	51	34	+17
44	rememberance	(noun)	53	43	+10
45	mischevious	(adjective)	44	38	+ 6
46	advantagous	(adjective)	72	58	+14
47	recieving	(verb)	39	36	+ 3
48	courtious	(adjective)	44	27	+17
49	mathmatics	(noun)	56	37	+19
50	aquaint	(verb)	47	37	+10
51	decend	(verb)	26	18	+ 8
52	prefered	(verb)	40	27	+13
53	raisen	(noun)	42	33	+ 9
54	signiture	(noun)	46	22	+24
55	totaly	(adverb)	44	19	+25
56	accomodation	(noun)	16	10	+ 6
57	alledged	(verb)	28	17	+11
58	satelite	(noun)	45	20	+25

[1] Any words that could be more than one part of speech were classified by the most common usage.

TABLE Adv-C: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART A: USAGE, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Verb agreement with subject	79	83	– 4
2	Irregular verb forms	83	84	– 1
3	Irregular verb forms	81	75	+ 6
4	Idiomatic use of preposition	72	87	–15
5	Word choice	40	67	–27
6	Verb tenses	48	63	–15
7	Irregular verb forms	68	86	–18
8	Possessive pronoun	63	63	0
9	Pronoun objective case	77	63	+14
10	Word choice involving verbs	61	65	– 4
11	Verb tenses	56	74	–18
12	Pronoun objective case	61	63	– 2
13	Pronoun objective case	71	73	– 2
14	Verb tenses	63	59	+ 4
15	Spelling	65	56	+ 9
16	Word choice involving verbs	40	39	+ 1
17	Word choice involving verbs	49	67	–18
18	Word choice involving verbs	25	41	–16
19	Pronoun objective case	39	47	– 8
20	Word choice	37	52	–15
21	Verb tenses	30	41	–11
22	Verb tenses	32	36	– 4
23	Possessive pronoun	66	70	– 4
24	Irregular verb forms	46	42	+ 4
25	Word choice involving verbs	48	37	+11
26	Double negative	28	47	–19
27	Verb agreement with subject	58	46	+12
28	Irregular verb forms	51	49	+ 2
29	Word choice	43	31	+12
30	Adjective – comparative, superlative forms	36	51	–15
31	Irregular verb forms	19	28	– 9
32	Spelling	42	34	+ 8
33	Pronoun nominative case	23	29	– 6
34	Substandard corruption form	43	17	+26
35	Substandard corruption form	18	18	0
36	Verb agreement with subject	42	20	+22
37	Pronoun objective case	38	14	+24
38	Word choice involving verbs: spelling	20	18	+ 2

TABLE Adv-D: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART B: PUNCTUATION, ADVANCED BATTERY, FORM W.

Item Number	Punctuation Required	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
39	Comma: Separating phrase	63	74	−11
40	Comma: Between city and state	89	77	+12
41	Comma: Separating phrase	52	67	−15
42	Comma: Separating phrase	49	58	− 9
43	Comma: Separating phrase	55	63	− 8
44	None required: in compound verb	53	63	−10
45	None required: between two adjectives	77	78	− 1
46	Comma: Separating participial phrase	46	58	−12
47	Quotation marks: Closing quotation in sentence	65	60	+ 5
48	Period: End of sentence	81	73	+ 8
49	Quotation marks: Closing quotation in sentence	61	63	− 2
50	Comma: Separating words in a series	63	79	−16
51	Optional: In a series, before "and"	27	20	+ 7
52	Quotation marks: Opening quotation in sentence	80	61	+19
53	Comma: Separating direct address	64	52	+12
54	Comma: Separating exclamation from sentence	70	57	+13
55	Quotation marks: Closing quotation within sentence	66	57	+ 9
56	Optional: Setting off phrase	24	16	+ 8
57	Quotation marks: Opening quotation within sentence	67	52	+15
58	Quotation marks: Closing quotation	67	54	+13

TABLE Adv-E: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART C: CAPITALIZATION, ADVANCED BATTERY, FORM W.

Item Number	Capitalization Rule	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
59	Name of a city	99	98	+ 1
60	No caps: common noun within sentence	84	87	− 3
61	Name of a school	99	98	+ 1
62	Name of a school	90	93	− 3
63	No caps: common noun within sentence	52	61	− 9
64	No caps: common noun within sentence	56	65	− 9
65	Religious event	67	83	−16
66	Religious event	93	91	+ 2
67	No caps: common noun within sentence	91	86	+ 5
68	No caps: common noun within sentence	83	81	+ 2
69	Person's title	54	75	−21
70	No caps: common noun within sentence	92	93	− 1
71	Title of institution	92	91	+ 1
72	No caps: preposition within title	91	81	+10
73	Title of institution	94	93	+ 1
74	No caps: common noun within sentence	90	86	+ 4
75	No caps: common noun within sentence	84	93	− 9
76	No caps: common noun within sentence	85	85	0
77	Name of a building	81	93	−12
78	Name of a street	88	90	− 2
79	No caps: common noun within sentence	94	93	+ 1
80	No caps: common noun within sentence	89	93	− 4
81	Name of a building	94	95	− 1
82	Institution	98	96	+ 2
83	Institution	96	95	+ 1
84	No caps: season of the year	93	94	− 1
85	No caps: common noun within sentence	86	79	+ 7
86	Title of organization	99	96	+ 3
87	Title of organization	99	96	+ 3
88	Title of organization	98	96	+ 2
89	Title of organization	95	96	− 1
90	Name of a person	81	81	0
91	Title of a composition	90	87	+ 3
92	No caps: preposition within title	94	88	+ 6
93	No caps: common noun within sentence	89	90	− 1
94	Institution	76	67	+ 9
95	No caps: common noun within sentence	39	40	− 1
96	Name of a language	93	86	+ 7
97	Nationality	97	94	+ 3
98	National institution	87	90	− 3
99	No caps: common noun within sentence	84	89	− 5
100	Literary title	99	95	+ 4
101	Literary title	95	89	+ 6
102	First word in complimentary close	94	91	+ 3
103	No caps: second word of a closing	74	75	− 1

TABLE Adv-F: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART D: DICTIONARY SKILLS, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
104	Choice of meaning	80	84	− 4
105	Choice of meaning	60	74	−14
106	Recognition of root word	60	71	−11
107	Guide words	67	57	+10
108	Choice of meaning	77	74	+ 3
109	Choice of meaning	40	64	−24
110	Accent mark	58	79	−21
111	Guide words	57	36	+21
112	Choice of meaning	58	81	−23
113	Choice of meaning	39	60	−21
114	Part of speech	60	54	+ 6
115	Accent mark	41	42	− 1
116	Choice of meaning	39	51	−12
117	Choice of meaning	59	46	+13
118	Spelling	42	41	+ 1
119	Choice of meaning	42	48	− 6
120	Pronunciation	78	89	−11
121	Pronunciation	53	70	−17
122	Pronunciation	55	71	−16
123	Pronunciation	54	61	− 7
124	Pronunciation	52	63	−11
125	Pronunciation	24	32	− 8
126	Pronunciation	17	45	−28
127	Pronunciation	31	44	−13

TABLE Adv-G: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, LANGUAGE SUB-TEST, PART E: SENTENCE SENSE, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
128	Incomplete sentence	69	88	−19
129	Incomplete sentence	53	79	−26
130	One complete sentence	81	81	0
131	Two complete sentences	66	64	+ 2
132	Incomplete sentence	58	56	+ 2
133	Two complete sentences	77	80	− 3
134	One complete sentence	72	84	−12
135	Incomplete sentence	64	45	+19
136	Two complete sentences	59	65	− 6
137	Incomplete sentence	22	67	−45
138	Two complete sentences	78	71	+ 7
139	Incomplete sentence	75	84	− 9
140	One complete sentence	56	73	−17
141	Two complete sentences	58	46	+12
142	Two complete sentences	71	63	+ 8
143	Incomplete sentence	46	70	−24
144	Two complete sentences	76	72	+ 4
145	One complete sentence	70	70	0

TABLE Adv-H: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, ARITHMETIC COMPUTATION SUB-TEST, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Division: 5d ÷ 2d [1]	67	75	− 8
2	Subtraction: 6d − 5d	82	79	+ 3
3	Division: 4d ÷ 2d	88	84	+ 4
4	Addition: Mixed numbers	73	80	− 7
5	Division: Decimals	74	77	− 3
6	Multiplication: 3d x 3d	70	68	+ 2
7	Subtraction: 6d − 5d	76	70	+ 6
8	Division: Decimals	88	86	+ 2
9	Subtraction: Fractions	72	78	− 6
10	Multiplication: Mixed numbers	69	81	−12
11	Addition: 4d + 4d + 4d + 4d	80	68	+12
12	Division: Mixed numbers and fractions	63	71	− 8
13	Rounding	46	50	− 4
14	Addition: Mixed numbers and fractions	72	76	− 4
15	Rounding	66	69	− 3
16	Multiplication: 3d x 3d	67	52	+15
17	Addition: 4d + 4d+ 4d + 4d	68	59	+ 9
18	Multiplication: 4d x 2d	53	47	+ 6
19	Subtraction: Mixed numbers	53	54	− 1
20	Percent	39	56	−17
21	Multiplication: Whole and mixed numbers	45	69	−24
22	Equation	61	53	+ 8
23	Rounding	53	48	+ 5
24	Percent	24	39	−15
25	Division: Decimals	54	45	+ 9
26	Percent	35	56	−21
27	Division: Decimals	51	49	+ 2
28	Equation	71	66	+ 5
29	Division: 2d ÷ 2d	33	40	− 7
30	Equation	46	39	+ 7
31	Equation	48	17	+31
32	Percent	20	23	− 3
33	Percent	24	33	− 9
34	Multiplication: Directed numbers	60	36	+24
35	Percent	23	22	+ 1
36	Equation	31	22	+ 9
37	Addition: Directed numbers	50	24	+26
38	Equation	38	13	+25
39	Division: Directed numbers	55	26	+29
40	Equation	29	7	+22
41	Exponent	35	13	+22

[1] The letter "d" stands for the number of digits in the computation, e.g., 5d ÷ 2d means 5 digits divided by 2 digits (45 $\overline{)13770}$).

TABLE Adv-I: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, ARITHMETIC CONCEPTS SUB-TEST, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Fraction concepts	67	83	−16
2	Rounding	57	78	−23
3	Roman numerals	79	81	− 2
4	Equation	90	77	+13
5	Number property and operational relationship	58	66	− 8
6	Fraction concepts	52	69	−17
7	Number property and operational relationship	63	69	− 6
8	Rounding	45	67	−22
9	Estimation	44	54	−10
10	Number property and operational relationship	77	60	+17
11	Percent	32	53	−21
12	Formula	51	51	0
13	Number property and operational relationship	55	69	−14
14	Fraction concepts	39	55	−16
15	Number property and operational relationship	35	57	−22
16	Number property and operational relationship	63	68	− 5
17	Roots	45	20	+25
18	Powers	73	32	+41
19	Directed number	67	36	+31
20	Measurement (time)	69	54	+15
21	Equation	53	45	+ 8
22	Number series	50	51	− 1
23	Number property and operational relationship	51	45	+ 6
24	Percent	43	41	+ 2
25	Number property and operational relationship	39	48	− 9
26	Estimation	46	40	+ 6
27	Vocabulary	64	61	+ 3
28	Number property and operational relationship	35	42	− 7
29	Number property and operational relationship	64	45	+19
30	Number property and operational relationship	33	32	+ 1
31	Fraction concepts	31	43	−12
32	Estimation	39	32	+ 7
33	Set	55	37	+18
34	Set	57	42	+15
35	Equation	29	29	0
36	Number property and operational relationship	24	22	+ 2
37	Prime number	48	31	+17
38	Non-decimal numbers	30	29	+ 1
39	Roots	7	13	− 6
40	Non-decimal numbers	19	15	+ 4

TABLE Adv-J: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, ARITHMETIC APPLICATIONS SUB-TEST, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Rate	69	80	−11
2	Measurement	68	62	+ 6
3	Reading a graph	41	52	−11
4	Reading a graph	63	57	+ 6
5	Reading a graph	45	55	−10
6	Reading a graph	63	62	+ 1
7	Reading a graph	42	43	− 1
8	Reading a graph	46	66	−20
9	Reading a graph	85	76	+ 9
10	Reading a graph	58	62	− 4
11	Measurement	49	47	+ 2
12	Proportion	43	49	− 6
13	Rate	64	73	− 9
14	Division problem	66	69	− 3
15	Problem analysis	67	78	−11
16	Problem analysis	13	16	− 3
17	Measurement	45	41	+ 4
18	Proportion	55	52	+ 3
19	Measurement	26	53	−27
20	Multiple-step problem	26	34	− 8
21	Problem analysis	42	41	+ 1
22	Multiple-step problem	16	21	− 5
23	Business arithmetic: Profit	14	20	− 6
24	Business arithmetic: Commission	23	33	−10
25	Business arithmetic: Discount	30	35	− 5
26	Reading a table	44	45	− 1
27	Reading a table	21	41	−20
28	Reading a table	31	26	+ 5
29	Rate	14	27	−13
30	Geometry	39	31	+ 8
31	Proportion	15	24	− 9
32	Geometry	21	17	+ 4
33	Business arithmetic: Interest	18	17	+ 1
34	Logical reasoning	13	18	− 5
35	Probability	8	12	− 4
36	Geometry	19	11	+ 8

TABLE Adv-K: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SOCIAL STUDIES SUB-TEST, PART A: CONTENT, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Sociology	72	92	−20
2	History	80	94	−14
3	Sociology	41	81	−40
4	Geography	58	82	−24
5	Civics	67	75	− 8
6	History	55	68	−13
7	Civics	59	66	− 7
8	Economics	48	68	−20
9	Civics	46	51	− 5
10	Vocation	74	80	− 6
11	History	68	74	− 6
12	Civics	52	63	−11
13	Civics	49	67	−18
14	Sociology	58	70	−12
15	Economics	54	61	− 7
16	Geography	63	68	− 5
17	Geography	61	55	+ 6
18	History	52	37	+15
19	Civics	64	65	− 1
20	Geography	55	55	0
21	Industry	28	40	−12
22	Economics	33	37	− 4
23	Geography	40	49	− 9
24	Geography	56	56	0
25	Geography	53	48	+ 5
26	Sociology	55	41	+14
27	Civics	42	46	− 4
28	Civics	38	57	−19
29	History	44	42	+ 2
30	Civics	31	37	− 6
31	Economics	47	49	− 2
32	History	32	35	− 3
33	Civics	49	38	+11
34	Sociology	38	47	− 9
35	History	18	46	−28
36	History	28	42	−14
37	Sociology	31	42	−11
38	Sociology	31	46	−15
39	Sociology	32	39	− 7
40	History	42	41	+ 1
41	Economics	25	39	−14
42	Geography	43	36	+ 7
43	Economics	32	31	+ 1
44	Sociology	43	38	+ 5
45	Civics	35	32	+ 3
46	History	34	37	− 3
47	Geography	23	26	− 3
48	Geography	11	25	−14
49	Economics	15	26	−11
50	Civics	22	26	− 4
51	Civics	16	14	+ 2
52	Civics	35	41	− 6

TABLE Adv-L: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SOCIAL STUDIES SUB-TEST, PART B: STUDY SKILLS, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
53	Reading a double bar or line graph	58	73	−15
54	Reading a double bar or line graph	60	69	− 9
55	Reading a double bar or line graph	52	75	−23
56	Reading a double bar or line graph	67	69	− 2
57	Reading a double bar or line graph	53	65	−12
58	Reading a double bar or line graph	49	55	− 6
59	Using references	46	69	−23
60	Using references	52	72	−20
61	Using references	62	67	− 5
62	Using references	45	62	−17
63	Using references	56	53	+ 3
64	Using references	45	35	+10
65	Using a bibliography	79	68	+11
66	Using a bibliography	60	63	− 3
67	Using a bibliography	61	47	+14
68	Using a bibliography	49	50	− 1
69	Using a bibliography	26	39	−13
70	Using a bibliography	38	34	+ 4
71	Using a library index card	54	66	−12
72	Using a library index card	65	77	−12
73	Using a library index card	38	58	−20
74	Using a library index card	55	65	−10
75	Using a library index card	27	30	− 3
76	Using a library index card	77	83	− 6
77	Reading a globe	86	79	+ 7
78	Reading a globe	53	55	− 2
79	Reading a globe	82	70	+12
80	Reading a globe	70	59	+11
81	Reading a globe	45	42	+ 3
82	Reading a globe	47	50	− 3
83	Reading a globe	56	46	+10
84	Reading a globe	63	45	+18
85	Reading a globe	64	45	+19
86	Reading a globe	39	37	+ 2
87	Reading a globe	41	35	+ 6
88	Interpreting a political poster	51	51	0
89	Interpreting a political poster	25	36	−11
90	Interpreting a political poster	25	29	− 4
91	Interpreting a political poster	39	32	+ 7
92	Interpreting a political poster	29	33	− 4

TABLE Adv-M: ITEM NUMBERS, CONTENT CLASSIFICATION, DIFFICULTY VALUES, AND DIFFERENCE VALUE, SCIENCE SUB-TEST, ADVANCED BATTERY, FORM W.

Item Number	Topic Measured	Item Difficulty Values		Difference
		Hearing Impaired Students	Hearing Students	
1	Plants	91	91	0
2	Conservation	77	90	−13
3	Scientific method	88	90	− 2
4	Electricity and magnetism	80	78	+ 2
5	Conservation	73	73	0
6	Energy and machines	54	63	− 9
7	Energy and machines	79	77	+ 2
8	Magnetism	75	81	− 6
9	Chemistry	73	73	0
10	The body	65	69	− 4
11	Earth science	60	64	− 4
12	Animals	61	76	−15
13	Plants and animals	47	61	−14
14	Scientific method	48	69	−21
15	The body	57	60	− 3
16	Earth science	54	76	−22
17	Science in industry	73	75	− 2
18	Astronomy	65	66	− 1
19	Chemistry	77	71	+ 6
20	The body	66	50	+16
21	Animals	51	66	−15
22	Weather; plants and animals	76	73	+ 3
23	Magnetism	42	61	−19
24	Plants	70	63	+ 7
25	The body	43	67	−24
26	The body	52	61	− 9
27	Sound	36	47	−11
28	Scientific term	49	44	+ 5
29	Earth science	69	51	+18
30	Air and weather	55	50	+ 5
31	Safety	64	61	+ 3
32	Animals	71	51	+20
33	Animals	59	52	+ 7
34	Electricity	45	52	− 7
35	Plants	43	41	+ 2
36	Heat	35	43	− 8
37	Plants	47	45	+ 2
38	Earth science	49	41	+ 8
39	Earth science	51	46	+ 5
40	Earth science	46	51	− 5
41	Astronomy	47	45	+ 2
42	Light	40	49	− 9
43	Astronomy	30	34	− 4
44	Plants	44	42	+ 2
45	Animals	36	37	− 1
46	Magnetism; scientific method	39	51	−12
47	Plants	37	35	+ 2
48	Earth science	25	26	− 1
49	Energy and machines	43	40	+ 3
50	Air and weather	33	38	− 5
51	Energy and machines	39	27	+12
52	Earth science	24	34	−10
53	The body	25	17	+ 8
54	Heat	28	27	+ 1
55	The body	20	21	− 1
56	Plants	17	18	− 1
57	Food	17	21	− 4
58	Energy	24	21	+ 3
59	Earth science	17	13	+ 4
60	Astronomy	21	18	+ 3

The Discriminative Validity of Selected Sub-tests of The Stanford Achievement Test, Intermediate I Battery, for Hearing Impaired Students: Spring 1971

Carol Buchanan

Questions of validity arise whenever a test is used for a population other than that for which it was standardized. The Stanford Achievement Test has been standardized for hearing students, and therefore its use with hearing impaired students necessitates an examination of its validity for this new group of individuals.

Assessments of validity evaluate the degree to which an instrument measures what it purports to measure: "Is it reasonable to assume that my students' true levels of achievement are being reflected by the scores from these tests?" The answer to this question involves multiple considerations. For achievement tests, validity is first of all a question of the adequacy with which the tests sample and represent the domains of knowledge or content contained in the school curriculum with which the student is involved. This, clearly, will vary from school to school and from class to class, so that there is no substitute for a careful analysis of the degree to which the test content matches that of the curriculum.

Assume that the content of the tests matches a given curriculum satisfactorily. The validity question now becomes one of how well the tests are doing their job of differentiating varying levels of achievement among the students in a specified group, for that is the purpose of "norm-referenced" tests such as the Stanford Achievement Test.[1] This characteristic of tests is evaluated by means of item discrimination coefficients, which reflect the degree to which the given item in fact successfully distinguishes between high and low achievers on the sub-test in question. An effective test item is answered correctly by a significantly higher proportion of high-achieving students than of low-achieving students. The discrimination coefficients measure the extent to which this desirable state of affairs actually occurs.

Item discrimination may be determined if the individual item performance of two groups representing high and low achievers is known. The item difficulties[2] of both groups lead to a discrimination coefficient which reflects the degree to which an item is able to differentiate the two groups and is an estimate of how well it discriminates among all students taking the test.

This report will provide information regarding the overall discrimination validity of selected subtests of the Intermediate I battery (Form W), as well as discrimination coefficients for each individual item. These indices are valuable as a means of determining the degree of differentiation, and the data on which they are based can frequently suggest explanations for particularly strong or weak performance of a given group on an item or a sub-test.

[1] This is in contrast to "content-referenced" or "criterion-referenced" tests in which the focus is on whether the individual student has or has not measured up to a pre-determined standard of achievement, rather than on differentiating among varying levels of achievement and, in effect, comparing students to each other.

[2] The "item difficulty" of an item is the proportion of test-takers who answer the item correctly, generally expressed as a proportion or percentage. Higher numbers indicate easier items, so this should probably be called an "item easiness" index, but the standard usage is retained here.

SELECTION OF THE BATTERY AND SUB-TESTS

The Intermediate I battery was selected for study because it was administered to more than 3,000 students, a significant proportion of all students tested in the Spring 1971 National Achievement Testing Program conducted by the Office of Demographic Studies. In addition, tests at this level were not modified for the hearing impaired and are therefore exactly the same for both hearing and hearing impaired students. (For a discussion of the modifications of the Primary batteries, see Appendix II of this publication.)

Five of the ten sub-tests were chosen for analysis. They were selected on the basis of their content and results obtained from the Achievement Testing Program in 1971. The shape of grade equivalent distributions, item difficulty values, and overall difficulty were all important factors in each sub-test's selection.

Paragraph Meaning

Paragraph Meaning is a test of reading comprehension. Scores students received on this sub-test were, for the most part, concentrated in the low to mid-portion of the intended range of scores. This is the result of the use of a Paragraph Meaning test as a screening device for all students in the National Testing Program. Screening rationale and mechanics are also described in full in Appendix II.

The average item was answered correctly by 40 percent of the students. The corresponding value is 53 percent for hearing students. Thus, hearing impaired students found this test more difficult. Many educators of the deaf agree that reading is one of the most important skills measured in the test series. It was included in this study on that basis, as well as on the basis that student performances were grouped homogenously below the ideal level and that this test was used as the instrument for screening.

Spelling

Many students taking the Spelling Sub-test scored above the norm for hearing students. Since this test presented little difficulty for many students, questions arose regarding the items' sensitivity in differentiating their spelling skills.

Arithmetic Computation

The distribution of scores obtained from the Arithmetic Computation test was similar to that described for Spelling. Hearing impaired students achieve best here, relative to other areas. The examination of item sensitivity, given the limitation that most students scored well, appeared necessary.

Arithmetic Applications

The selection of Arithmetic Applications was based more upon the nature of its content than on observations regarding performance. The eight percent difference between the average item difficulties of the hearing impaired and hearing groups is a significant but not overwhelming difference.

The items in this test are word problems. Hearing impaired students scored well in the Computation Sub-test where arithmetic operations had to be performed, but this sub-test presented more of a challenge in that the students must process verbal information in order to arrive at the mathematical form for solution.

Science

The items comprising this sub-test were chosen by the test authors on the basis of a survey of the topics contained in science curriculums for average 4th and 5th grade hearing students in regular classes. This could present limitations in content validity for hearing impaired students, which may or may not be reflected in their performance. The average item difficulty was .46, ten units below the difficulty for hearing students. Item discrimination coefficients here may be of some aid to those interested in exploring differences between hearing and hearing impaired students in a "subject matter" area as opposed to a "basic skill" area.

SELECTION OF STUDENT SAMPLES

As mentioned previously, item discriminations in this study are determined through the analysis of the individual item performance of two contrasting groups of students, henceforth referred to as the Upper and Lower level students. These groups are random samples of the 27 percent highest and lowest achievers within a particular sub-test. These students were administered the Intermediate I battery as a part of the 1971 Achievement Testing Program. Table 1 shows the grade equivalent scores which bounded the 27th percentiles at each end of the sub-tests' distributions.

TABLE 1: THE GRADE EQUIVALENT SCORES WHICH DETERMINED THE UPPER AND LOWER 27TH PERCENTILES FOR EACH SUB-TEST.

Categories	Paragraph Meaning	Spelling	Arithmetic Computation	Arithmetic Applications	Science
Grade Equivalent Scores Which UPPER Level Students Scored At or ABOVE	4.3	6.8	6.7	5.2	4.6
Grade Equivalent Scores Which LOWER Level Students Scored At or BELOW	3.1	4.6	4.8	3.8	3.7

As over 3,200 students were administered the battery, equal amounts of approximately 860 students comprised each of the Upper and Lower levels. All of these students were identified. Ten samples of approximately 35 students were randomly selected to represent each Upper and Lower group of the five sub-tests. Each sample's representativeness of its total group of Upper and Lower students is reported in Table 2 in terms of average grade equivalent scores. The greatest difference in scores between the samples and their total groups exists in the Upper level of the Science Sub-test. They differ by less than one-tenth of a year.

The 27th percentiles are traditionally used in item discrimination studies for the determination of Upper and Lower students. This provides data giving the closest estimate of the discrimination coefficient for analyses of a very large group. As the number of subjects decreases, the best approximations can be obtained by using cut-offs which approach the 33rd percentiles. Analyses on a classroom of students, for example, would probably be conducted using the upper and lower 33 percent of students.

PROCEDURES IN DETERMINING ITEM DISCRIMINATION

When the students were tested in 1971, they marked their answers to each item on an answer sheet. These answer sheets were located from storage files for all the students in the ten samples. The choices each student indicated for each item in the sub-test were transcribed onto data sheets.

A percent of correct responses by the Upper and Lower groups was calculated for each item. If, for example, the number of Upper students correctly answering item No. 11 of Paragraph Meaning was 24 out of 32 responses, the percent (U) would equal 75. A corresponding percent (L) was calculated from the Lower group's responses on that same item. When students omitted an answer or chose more than one answer for an item, their responses (or non-response)

TABLE 2: COMPARISON OF THE AVERAGE GRADE EQUIVALENT SCORES OF THE LOWER AND UPPER GROUP SAMPLES WITH THOSE OF ALL STUDENTS* SCORING IN THE UPPER AND LOWER 27 PERCENTILES.

	Paragraph Meaning		Spelling		Arithmetic Computation		Arithmetic Applications		Science	
Average Grade Equivalent Scores	Sample	Total Group of Students	Sample	Total Group of Students	Sample	Total Group of Students	Sample	Total Group of Students	Sample	Total Group of Students
Lower 27 Percent	2.80	2.82	4.03	4.02	3.99	4.03	3.28	3.29	3.44	3.42
Upper 27 Percent	4.83	4.87	7.87	7.79	7.98	8.01	6.18	6.17	5.32	5.23

*The approximately 3,200 students administered each sub-test of the Intermediate I battery in the 1970-71 National Achievement Testing Program.

were included in the calculations as a wrong answer. In cases where it appeared a response was not indicated because time did not permit students to complete the test, this was *not* regarded as a wrong answer and was excluded from the calculations. Omitted items and items which were not answered because of an apparent time lapse were distinguished by their location. If answers were chosen for the items which directly preceded and followed a non-response, the item was classified as having been omitted. If the item was part of a continuous span of non-responses through the end of the test, it was classified as being unanswered due to a lapse in the time allowed to complete the test.

Each pair of percents was converted into a product-moment correlation coefficient. Flanagan's (1939) table[1] was used in making the conversions. These values estimate the correlation between performance on the item and overall knowledge of the subject area as measured by the sub-test. If an item is effective in differentiating achievement levels, the Upper group, being the higher achievers, should have a higher proportion of correct answers than the Lower group. The significance of these differences in passage rates is reflected in the magnitude of the correlation. The higher the discrimination coefficient, the better an item is differentiating achievement levels. These values have the theoretical range of −1.00 to +1.00.

RESULTS

Table 3 reports the Upper and Lower difficulty levels and discrimination coefficient for each item of the five sub-tests. The highest difficulty levels (i.e., the easiest items) appear towards the beginning of a sub-test. The summary of the average discrimination values for each sub-test shows the items of Spelling, Arithmetic Computation, and Arithmetic Applications to have the highest average differentiating levels. The first two of these are the sub-tests on which hearing impaired students performed highest. All three sub-tests also produced the greatest differences between the average grade equivalent scores of their Upper and Lower groups, as shown in Table 4.

To facilitate the reporting and analysis of the results, two points within the range of item discriminations were chosen to differentiate high and low discriminating items. These two points were chosen arbitrarily but can be helpful in interpreting results. Correlations less than or equal to .15 have been classified as low discriminations. Those at or above .70 are considered above average. The middle range consists of coefficients ranging from .16 to .69. Table 5 presents the proportions of items in each sub-test that fall into the defined intervals.

In each of the sub-tests, most of the items fall within the middle range of discrimination. Arithmetic Computation has the greatest percentage of items discriminating "highly." Paragraph Meaning has the highest percentage of "low" discriminating items.

Table 6 shows for the three classes of items in Table 5 the corresponding item difficulties obtained from all students in the 1971 testing program. From this table, the trend for four of the sub-tests appears to be that the items which discriminate best are the easier items. Arithmetic Computation shows the reverse tendency, with the high discriminating items being the more difficult ones.

THE GUESSING FACTOR AS IT RELATES TO THE PERFORMANCE OF THE LOWER GROUP

When dealing with scores at the lower end of a distribution, consideration must be given to the possibility of student guessing. Every multiple-choice test has a score which can be obtained if a student marks his answers randomly, without regard to the item question or answer choices. This score is the "chance score" and is easily derived. If, for example, an item has four response options, the probability is one out of four that a pure guess will be right. If this probability is multiplied by the total number of items in the test, a score will be obtained which represents the expected number of correct answers for the entire test.

A portion of students in each Lower level scored at or below the chance level. These students may have guessed or they may have made an earnest attempt to choose correct answers. If there was a great deal of guessing among Lower level students, this would understandably create limitations in the validity of the results in this report. It was decided, therefore, to determine the extent of guessing since the chance score is within the range of scores for all the Lower groups.

It is not possible to determine with certainty whether a student guesses unless he is asked his reaction to each item. One alternative is to compute an individual chance score for each student who

[1] Flanagan, J.C. "General considerations in the selection of test items and a short method of estimating the product-moment coefficient from the data at the tails of the distributions," *Journal of Educational Psychology*, 1939, 30, 674-680.

TABLE 3: DISCRIMINATION COEFFICIENTS AND THE PERCENTAGE OF CORRECT ITEM RESPONSES* FOR THE UPPER AND LOWER 27% OF STUDENTS IN THE DISCRIMINATION STUDY FOR SELECTED SUB-TESTS OF THE INTERMEDIATE I BATTERY.

Sub-test Item Numbers	Selected Sub-tests of the Intermediate I Battery														
	Paragraph Meaning			Spelling			Arithmetic Computation			Arithmetic Applications			Science		
	Percent Correct			Percent Correct			Percent Correct			Percent Correct			Percent Correct		
	Upper 27%	Lower 27%	r**	Upper 27%	Lower 27%	r**	Upper 27%	Lower 27%	r**	Upper 27%	Lower 27%	r**	Upper 27%	Lower 27%	r**
1	62	44	.18	91	74	.28	94	94	.00	94	75	.34	100	68	.63+
2	88	11	.74	97	39	.70	91	60	.42	85	31	.55	100	68	.63+
3	35	3	.54	100	45	.74+	100	77	.56+	74	16	.58	94	62	.47
4	71	25	.46	97	65	.54	94	83	.24	74	19	.55	94	41	.62
5	24	11	.21	97	81	.40	97	89	.26	59	6	.62	82	30	.53
6	68	25	.44	100	74	.59+	94	77	.32	68	9	.62	88	49	.46
7	53	11	.49	100	71	.62+	91	38	.58	100	78	.55+	53	43	.10
8	97	67	.53	100	77	.56+	100	74	.59+	94	38	.64	76	22	.54
9	94	36	.65	100	58	.69+	97	29	.75	53	22	.34	94	41	.62
10	88	31	.59	97	61	.58	100	65	.66+	65	22	.44	24	19	.07
11	62	33	.30	100	90	.40+	100	32	.80+	97	19	.80	88	49	.46
12	41	33	.09	100	71	.62+	100	53	.72+	82	25	.57	68	43	.26
13	53	42	.11	100	55	.70+	79	62	.20	82	9	.72	88	51	.44
14	74	22	.52	100	45	.74+	97	68	.52	71	0	.82+	91	62	.40
15	68	36	.33	94	68	.42	97	59	.59	88	22	.66	91	27	.66
16	82	25	.57	88	45	.48	88	50	.45	94	53	.54	74	38	.37
17	62	39	.24	100	39	.78+	97	24	.78	97	19	.80	100	57	.70+
18	56	39	.17	100	77	.56+	97	39	.70	79	9	.70	91	49	.51
19	100	58	.69+	97	52	.64	91	24	.68	82	22	.60	82	38	.47
20	62	22	.42	100	61	.68+	91	18	.72	68	9	.62	88	24	.64
21	50	14	.42	97	71	.50	100	30	.81+	9	25	−.27	82	22	.60
22	65	22	.44	97	42	.69	100	30	.81+	79	16	.62	71	54	.18
23	62	19	.46	100	74	.59+	97	22	.78	74	19	.55	47	24	.25
24	68	25	.44	100	61	.68+	94	26	.71	91	6	.82	41	41	.00
25	74	36	.39	88	19	.68	91	32	.62	76	19	.57	76	27	.49
26	26	8	.30	91	68	.34	94	19	.75	47	38	.10	79	30	.50
27	50	25	.27	100	6	.90+	85	47	.42	71	6	.69	79	27	.52
28	71	61	.11	100	55	.70+	94	21	.74	79	12	.66	85	24	.61
29	53	28	.26	97	68	.52	97	24	.78	50	3	.64	76	24	.52
30	38	17	.26	94	48	.57	82	14	.67	50	12	.45	65	27	.39

*Percentage of correct item responses is equal to the number of correct responses divided by the total number of responses.

**"r" = the discrimination coefficient. A "+" sign following a coefficient indicates cases in which the Upper Group had a 100% correct figure or in which the Lower Group had a 0% correct figure.

TABLE 3 (continued): DISCRIMINATION COEFFICIENTS AND THE PERCENTAGE OF CORRECT ITEM RESPONSES FOR THE UPPER AND LOWER 27% OF STUDENTS IN THE DISCRIMINATION STUDY FOR SELECTED SUB-TESTS OF THE INTERMEDIATE I BATTERY.

Sub-test Item Numbers	Selected Sub-tests of the Intermediate I Battery														
	Paragraph Meaning			Spelling			Arithmetic Computation			Arithmetic Applications			Science		
	Percent Correct			Percent Correct			Percent Correct			Percent Correct			Percent Correct		
	Upper 27%	Lower 27%	r	Upper 27%	Lower 27%	r	Upper 27%	Lower 27%	r	Upper 27%	Lower 27%	r	Upper 27%	Lower 27%	r
31	88	47	.47	82	42	.43	88	11	.74	62	29	.34	71	32	.39
32	85	53	.38	73	35	.39	94	35	.66	47	16	.36	44	11	.41
33	91	25	.67	88	45	.48	97	32	.74	44	26	.20	65	38	.28
34	94	40	.62	70	39	.32	97	32	.74				56	19	.40
35	65	20	.46	100	39	.78+	91	30	.64				47	38	.10
36	56	17	.42	88	29	.60	94	13	.79				24	16	.12
37	47	40	.07	100	39	.78+	85	35	.52				71	32	.39
38	24	31	−.09	97	39	.70	76	17	.59				41	11	.39
39	74	23	.51	85	26	.59	82	26	.56				35	27	.09
40	32	20	.15	97	17	.81							62	59	.03
41	32	24	.10	100	37	.78+							59	16	.46
42	59	29	.31	88	23	.65							26	19	.10
43	15	9	.12	100	37	.78+							24	14	.15
44	50	39	.12	97	33	.73							38	14	.31
45	38	19	.24	100	17	.86+							97	54	.62
46	24	12	.19	73	17	.56							59	17	.45
47	74	32	.42	52	20	.35							19	17	.03
48	53	26	.29	97	20	.79							38	11	.36
49	35	39	−.04	55	53	.02							22	18	.06
50	41	19	.26	82	23	.59							39	15	.30
51	74	27	.47										61	30	.32
52	38	13	.32										23	42	−.22
53	44	57	−.13										27	6	.37
54	58	14	.48										50	19	.34
55	6	11	−.13										66	12	.57
56	23	21	.03										17	19	−.03
57	32	7	.40												
58	19	4	.34												
59	13	16	−.05												
60	45	32	.14												
$\bar{X}u$*	55.43			92.92			93.15			72.27			63.36		
$\bar{X}l$**	27.23			47.80			41.31			22.12			31.91		
$\bar{X}r$***	.319			.600			.605			.556			.365		

*Mean percent, Upper Group

**Mean percent, Lower Group

***Mean discrimination coefficient

TABLE 4: DIFFERENCES BETWEEN THE AVERAGE GRADE EQUIVALENT SCORES OF THE UPPER AND LOWER GROUPS OF THE ITEM DISCRIMINATION STUDY SAMPLES.

Categories	Paragraph Meaning	Spelling	Arithmetic Computation	Arithmetic Applications	Science
Upper Group	4.83	7.87	7.98	6.18	5.32
Lower Group	2.80	4.03	3.99	3.28	3.44
Difference Between the Two Groups	2.03	3.84	3.99	3.90	1.88

TABLE 5: GROUPED FREQUENCY AND PERCENTAGE DISTRIBUTIONS OF ITEM DISCRIMINATION COEFFICIENTS.

Item Discrimination Coefficients	Paragraph Meaning		Spelling		Arithmetic Computation		Arithmetic Applications		Science	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All Items	60	100.0	50	100.0	39	100.0	33	100.0	56	100.0
$r \leqslant .15$	15	25.0	1	2.0	1	2.5	2	6.1	13	23.2
$.15 < r < .70$	43	71.7	30	60.0	21	53.9	25	75.7	42	75.0
$r \geqslant .70$	2	3.3	19	38.0	17	43.6	6	18.2	1	1.8

TABLE 6: AVERAGE ITEM DIFFICULTY VALUES* FOR EACH SUB-TEST, ACCORDING TO ITEM DISCRIMINATION COEFFICIENT.

Item Discrimination Coefficients	Paragraph Meaning	Spelling	Arithmetic Computation	Arithmetic Applications	Science
All Items	40.2	73.0	69.5	46.2	46.3
$r \leqslant .15$	31.3	46.0	94.0	30.0	28.2
$.15 < r < .70$	42.3	73.5	71.0	47.0	51.3
$r \geqslant .70$	63.5	73.7	66.1	48.5	72.0

*Obtained from all students in the 1971 Achievement Testing Program.

scores at or below the guessing level. The theoretical chance score is computed for a sub-test based on the assumption that students complete every item. Some students omit items or do not complete a test. If the chance score for a 60 item test is 15 and a student attempted only ten items, all of his answers being correct, his score is within the test's theoretical guessing range. However, it is highly unlikely that he could have gotten all ten correct if he had guessed or randomly marked his answers. A new chance score can be computed based on the number of items he attempted rather than on all of a test's items.

Table 7 shows the percent of students in the Lower group who scored at or below the theoretical chance score for each test. The second column shows the adjusted percent based on the number of items each student attempted.

All sub-tests except Arithmetic Applications showed a decrease in the percent of chance-level students when consideration was given to the number of items each attempted. These percents should not be interpreted to reflect the actual percent of students who guessed, since there is no way of knowing for certain. They can be best interpreted as

TABLE 7: PERCENT OF STUDENTS IN THE LOWER GROUP WHOSE SCORES FELL AT OR BELOW THE THEORETICAL CHANCE LEVEL COMPARED TO THE ACTUAL PERCENT OF CHANCE SCORES WITH RELATION TO THE NUMBER OF ITEMS ATTEMPTED.

Sub-tests	Percent of Students At or Below the Theoretical Chance Level	Percent of Students At or Below Chance Based on the Number of Attempted Items
Paragraph Meaning	44.4	25.0
Spelling	6.4	3.2
Arithmetic Computation	11.4	2.8
Arithmetic Applications	37.5	37.5
Science	16.2	10.8

the maximum percent probably guessing. Twenty-five percent of Paragraph Meaning's Lower group scored at or below the chance level. The percent who actually did guess can be theoretically anywhere from 0 to 25 percent.

DISCUSSION

The data in this report are the results of a study of item discrimination within five sub-tests of the Stanford Intermediate I battery. Sub-tests which produced similar achievement levels [refer to page 23 of publication D-9, *Academic Achievement Test Results of a National Testing Program for Hearing Impaired Students*] also produced similar average discrimination coefficients, showing the relationship between the dispersion of student achievement levels and the degree of item discrimination.

Discrimination values varied from −.28 to +.99+. Fewer were found at the lower extreme. Coefficients of .99+ were the maximum values that could be obtained. The table which provided these values made no provision for either group's item difficulty being 0 or 100 percent, since in very large samples these absolute values are virtually non-existent. A + sign was added to coefficients to indicate those derived from cases in which either a 0 or 100 percent difficulty value occurred.

Discrimination coefficients can be used to locate inefficient items for special attention. Examination of the distributions of the response choices made by the Upper and Lower groups can often be helpful in generating explanations for the low discrimination efficiency of an item. Imprecise wording in a distractor (incorrect response option) or in the stem or correct answer choice may have misled the Upper group. Approximately equal rates of endorsement of all available response options by both groups might suggest that the topic of the item is not part of the students' curriculum or general fund of knowledge. Such explanations are, of course, speculative and should be accepted or rejected on the basis of further evidence. They cannot be considered "proven" on the basis of the original evidence obtained in the discrimination study.

Examples of items which represent the two extreme levels of discrimination ($r \leqslant .15, r \geqslant .70$) are presented in the discussion of the results for each sub-test. Below are hypothetical examples of data which would characterize low and high discriminating items. The examples include response rates for each answer choice. The values for the correct answer are enclosed in bold lines.

High Discriminating Item:

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	--	2	98	--
Lower group	3	53	42	2

Discrimination Coefficient = .72

Most of the Lower group chose alternative "b." That alternative is misleading primarily to students who do not achieve well, since it only misled 2 percent of the Upper group. The correct answer was chosen by 98 percent of the Upper group and 42 percent of the Lower group. This example shows that an item does not have to be difficult for the Lower group in order to discriminate well.

Low Discriminating Item:

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	54	20	16	10
Lower group	50	24	17	9

Discrimination Coefficient = .04

Items such as this, with similar patterns of response rates for both groups, contribute very little to the differentiation between high and low achieving students. The low discrimination coefficient reflects this inefficiency.

Considerations other than discriminative efficiency enter into the decision as to whether or not to use a given item in a test. For example, an item that is much too easy to discriminate well may nevertheless be included near the beginning of the test as a "warm-up" item. Total test scores are not adversely affected if the number of such low-discriminating items is minimal relative to the length of the test.

Paragraph Meaning Sub-test

The Paragraph Meaning Sub-test provides a measure of student ability in reading comprehension. The questions consist of sentences in connected discourse to form paragraphs. The student reads the paragraphs and chooses from four options the alternative that best fits the meaning of the paragraph or best answers questions about it. The vocabulary level of the paragraphs is intended to be sufficiently low so the test does not become a measure of word knowledge, but rather one of understanding and drawing proper conclusions from related sentences.

The use of a Paragraph Meaning test for purposes of screening affected the distribution of the Paragraph Meaning scores obtained in the National Testing Program, resulting in a narrow score spread. The resulting slight score differences among students in Paragraph Meaning reduced the ability of individual items to discriminate differences among students. Had the scores been spread more, the test items would have had more opportunity to discriminate differences.

The low discrimination correlations obtained reflected this restricted range of scores. There is a correction factor[1] for such a situation that is used to estimate what a correlation would be, had its range not been restricted. The individual item discrimination coefficients and their average reported in Table 3 for Paragraph Meaning are the actual obtained values. The average coefficient is .32, below the desired minimum of .40. When the correction factor was applied to this value, the adjusted average coefficient was approximately .46.

A proper evaluation of a test's item discrimination ability is made when the subjects are of the ability levels for which the test was designed. The correction factor simulates a less restricted range of scores, and is, therefore, a fairer and more reliable indication of the strength of the items' discriminating ability.

The adjusted average discrimination coefficient is now within the desired range and is comparable to the values for Spelling and Arithmetic Computation. An adjustment has not been made for the individual item discriminations; thus, the percent of low discriminating items (25 percent) reported in Table 5 for Paragraph Meaning is greater than it would be with the adjustment (17 percent).

The average grade equivalent obtained by the 2,817 students who took this sub-test in the National Testing Program was 3.8. The average item difficulty of 40 percent is acceptable, but indicated that the test is relatively difficult for hearing impaired students. These figures are, however, well above the "chance" or "guessing" level (25 percent average item difficulty and an average grade equivalent of 2.8), and are therefore within the range of accurate measurement of the Intermediate I battery. The following are examples of first the high, then the low discriminating items of this sub-test.

Test item No. 2 has a discrimination coefficient of .74. Note that 88 percent of the students in the Upper group correctly answered it while 11 percent of those in the Lower group were able to do so. Response percentages for each group on each answer choice are shown below.

[1] $$R_{12} = \frac{r_{12}\frac{S_1}{S_1}}{\sqrt{1 - r_{12}^2 + r_{12}^2\frac{S_1^2}{S_1^2}}}$$

,where R_{12} = the corrected coefficient, r_{12} = the original, uncorrected coefficient, S_1 = the standard deviation of the scores of the restricted group, and S_1 = the standard deviation of the scores of the unrestricted group.

The enemy guards were not watching. Bob decided that this was the time to __2__ .

2. a **escape** c **join**
b **capture** d **sleep**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	88	6	--	6
Lower group	11	17	8	64

Discrimination Coefficient = .74

Item No. 19 obtained a discrimination coefficient of .69+. Everyone of the Upper group students answered it correctly while 58 percent of the Lower group did so.

Although quite large, tortoises do not equal the size of turtles, which, keeping entirely to the sea for a living, have grown __18__ in the midst of unfailing plenty. Since the forelimbs of turtles, as is true of sea mammals, have been converted into paddles, they are good __19__ .

19. a **flyers** c **sleepers**
b **runners** d **swimmers**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	--	--	--	100
Lower group	11	19	11	58

Discrimination Coefficient = .69+

Item No. 53 discriminated inversely, with the Upper group answering correctly less often than the Lower group.

The name Low Countries was formerly applied to the Netherlands or Holland (Hollow Land), of which Belgium once formed a part. It indicates the low-lying nature of the land, which is chiefly a delta formed by the Rhine, Meuse, Waal, and Schelde Rivers, which flow through it into the North Sea. Before England was separated from the continent, nearly all Holland and part of the east of England were under water, making a great bay of the North Sea. As the sea gradually retreated northward, the Rhine flowed as far as Cromer to meet it. The hand of man has pushed the sea yet farther north, till now a quarter of the whole of Holland is below what was high-tide level at Amsterdam before the sea was banked up; another eighth is less than 40 inches higher. The sea has been pushed back by the building of dikes. Dikes, too, enclose the rivers and lakes and are used to reclaim marshy lands and gradually to turn them into fertile lands.

How many rivers are mentioned by name?

53. a **one** c **three**
b **four** d **two**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	15	44	29	12
Lower group	11	57	25	4

Discrimination Coefficient = –.13

Item No. 5 does not appear to be appropriate for hearing impaired students, in that the question involves the sound of a voice. Its discrimination does not fall below .15, but the distribution of responses suggests that the Upper group may have guessed the item ($X^2 = 0.3529$, d.f. = 3, $.90 < p < .95$).

Although the girls were not related, their resemblance was striking. Even their voices sounded __5__ .

5. a **different** c **high**
b **alike** d **sharp**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	24	24	29	24
Lower group	22	11	47	19

Discrimination Coefficient = .21

Spelling Sub-test

The Spelling Sub-test consists of 50 multiple-choice items, each containing four words. The student chooses from the four words the one that is spelled incorrectly. Research results have shown[1] that

[1] Kelley, T. L., *et al. Stanford Achievement Test Technical Supplement.* New York: Harcourt, Brace & World, Inc., 1966, p. 29.

the ability to recognize incorrectly spelled words correlates very highly (.86 – .90) with the ability to write correctly spelled words in a dictation type spelling test. Nearly all the words used are within the first 5,000 words in children's usage, and therefore the extent of a child's vocabulary should be relatively independent of his spelling ability. While these assertions have not been specifically tested for hearing impaired children, spelling is, on the average, one of the best areas of performance on achievement tests for them.

The results show that this test is quite efficient in differentiating levels of spelling achievement, with an average item discrimination coefficient of .60. The average item was answered correctly by 93 percent of the Upper group and 48 percent of the Lower group. Both figures are high for their respective groups, confirming the high performance typically observed on this test for hearing impaired students compared to other sub-tests of the same battery. Despite its easiness for some students, the items of this sub-test are sensitive in distinguishing levels of spelling skill.

All discrimination coefficients in the test, except one, ranged from .28 to .99+. Nineteen of the 50 items discriminated very well ($r \geqslant .70$). The overall distribution of the item discrimination coefficients reflects the soundness and sensitivity of the items of the Spelling Sub-test.

Items No. 27 and No. 40 are examples of the many items with high discrimination coefficients. One hundred percent of the Upper group and six percent of the Lower group answered No. 27 correctly. The fact that all Upper level students passed the item enhances the value of its discrimination coefficient of .90.

27. a **merchants** c **sword**
b **tardy** d **colection**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	–	–	–	100
Lower group	19	32	42	6

Discrimination Coefficient = .90+

The discrimination value for No. 40 is .81, with 97 percent and 17 percent of the Upper and Lower groups, respectively, passing the item.

40. a **discovery** c **possable**
b **haunted** d **signal**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	–	3	97	–
Lower group	10	63	17	10

Discrimination Coefficient = .81

The only item that was not able to discriminate well was No. 49. The percents of Upper and Lower group students passing the item are 55 percent and 53 percent, respectively. Alternatives a and c were chosen most frequently by both groups.

49. a **depot** c **foriegn**
b **boundary** d **auditorium**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	42	3	55	–
Lower group	30	13	53	3

Discrimination Coefficient = .02

Arithmetic Computation Sub-test

Arithmetic Computation tests a student's computational skills in addition, subtraction, multiplication, division, and the fractional part of a whole. There are 39 multiple-choice items in which the student may choose from four alternatives the correct answer to an incomplete operation. A fifth alternative available to students is that the correct answer is "not given" among the four choices.

This sub-test is quite similar to Spelling in the criteria used for its selection in this study and the results it produced concerning discrimination validity. Arithmetic Computation is an area in which hearing impaired students score above the standardized norm. It was found, despite its easiness, to discriminate well. Forty-four percent of the items reported high discriminations, and only one discriminated poorly. The average item discrimination value of .61 is quite satisfactory. No weaknesses in content or format were found.

Below are two items with high discrimination coefficients.

9. $\begin{array}{r} 817 \\ -749 \\ \hline \end{array}$

a **68**
b **77**
c **78**
d **66**
e **NG**

	Percent of Students Choosing Each Alternative				
	a	b	c	d	e
Upper group	97	--	--	--	3
Lower group	29	6	12	--	47

Discrimination Coefficient = .75

26.

$\frac{1}{9}$ **of 414 =** a **43, rem 2**
b **45**
c **46**
d **51 5/9**
e **NG**

	Percent of Students Choosing Each Alternative				
	a	b	c	d	e
Upper group	--	3	94	--	3
Lower group	13	6	19	19	26

Discrimination Coefficient = .75

The item below is the only addition problem in the test for which no carrying is necessary. It is also the only item with a discrimination coefficient at or below .15. The item content is at too easy a level to discriminate well, its purpose being mainly a warm-up for students as they are beginning the test.

1. **32**
+86

a **114**
b **118**
c **124**
d **128**
e **NG**

	Percent of Students Choosing Each Alternative				
	a	b	c	d	e
Upper group	--	94	--	3	3
Lower group	3	94	--	3	--

Discrimination Coefficient = .00

Arithmetic Applications Sub-test

The Arithmetic Applications Sub-test consists of 33 multiple choice items which measure reasoning with problems taken from life experiences. The general reading vocabulary as determined for normally hearing students has been kept much below the problem-solving level being measured. Computation difficulty has been controlled so that it is only a minor factor.

The student is required to apply his mathematical knowledge and ability to think mathematically in practical situations which concern area, volume, ratio, graphs, tables, scales, percent, business transactions, averages, problems with circles and other geometric figures, and the selection of mathematical models for problems.

The average item discrimination coefficient for this sub-test is .54. Of all items, only two discriminated poorly, and these were of identical structure. The remainder of items (94 percent) were in a desirable range of discrimination coefficients. The average item was answered correctly by 46 percent of all students in the spring testing program. This is a reliable index of the overall validity of the sub-test for use with hearing impaired students.

Items No. 11 and No. 18 are examples of the items in this sub-test having high discrimination coefficients. The former measures the recognition and use of a ratio in solving a problem.

11. A map reads, "1 inch = 100 miles." How far is 3 inches on the map?

a **4 mi.** b **100 mi.** c **102 mi.** d **103 mi.** e **NG**

	Percent of Students Choosing each Alternative				
	a	b	c	d	e
Upper group	--	--	--	3	97
Lower group	13	9	9	50	19

Discrimination Coefficient = .80

Item No. 18 shows 79 percent of the Upper group and 9 percent of the Lower group answering correctly, yielding a .70 discrimination coefficient. This item and No. 11 reflect a tendency for low achievers to favor addition over any other operation as a means of solving a problem.

Several Boy Scouts are going on a hike. Some of them are planning for it. Here are some of their problems.

18. How much more will the wieners cost than the buns? The wieners cost $2.40 and the buns cost 40 cents.

a **$1.00** b **$2.00** c **$1.30** d **$2.80** e **NG**

	Percent of Students Choosing Each Alternative				
	a	b	c	d	e
Upper group	3	79	--	18	--
Lower group	--	9	9	75	6

Discrimination Coefficient = .70

The two items discriminating poorly are presented below and are the only items of this type. Both groups performed much better on No. 26. However, this difference in the passage rates of such similar items is perhaps attributable to the answer in No. 21 being "not given." The first report in this publication comparing performance patterns of normally hearing and hearing impaired students gives data supporting a tendency for hearing impaired students to perform poorer on items for which the correct answer is negative (e.g., "not given").

21. Six scouts are going on this hike. Bob says this is only 1/3 of their troop. How many are in the troop?

a **2** b **3** c **9** d **24** e **NG**

	Percent of Students Choosing Each Alternative				
	a	b	c	d	e
Upper group	68	18	6	--	9
Lower group	19	19	22	9	25

Discrimination Coefficient = –.27

26. On pet day, Tom brought 3 rabbits. "That's 1/4 of mine," he said. How many rabbits has Tom in all?

a **5/4** b **3** c **18** d **12** e **NG**

	Percent of Students Choosing Each Alternative				
	a	b	c	d	e
Upper group	3	6	--	47	44
Lower group	12	12	16	38	19

Discrimination Coefficient = .10

The figures from these two items seem to indicate that the poor discrimination is a result of difficulty; that is, hearing impaired students at the Intermediate I level are apparently unable to determine a number which represents a whole from a number whose fractional portion is known. Difficult items are helpful in discriminating among the better students and should not be removed from a test when their incidence is low. The only significant weakness in this test occurs in the structure of No. 21, which discriminated negatively. The elements in its structure responsible for this are alternatives "a" and "e," which served to mislead the Upper group to the degree that they performed less well than the Lower group.

Science Sub-test

This sub-test is intended to measure the following:

1) the ability to see the application of the principles of science in our environment and everyday activities;
2) knowledge of the facts, vocabulary, and generalizations from the various branches of the natural sciences;
3) some knowledge of the scientific method.

Students are to choose the best of four answers in order to complete a given statement or to answer the question posed in the item.

Science is similar to Paragraph Meaning in its difficulty for hearing impaired students. The average discrimination of all items is .36. Thirteen of the 56 items discriminated poorly. These 13 are characterized by their difficulty, as reported in Table 6. The average item with poor discrimination was answered correctly by only 28 percent of all students in the 1971 testing program. The average item with moderate discrimination ($.15 < r < .70$) was correctly answered by 51 percent of students, with the one item discriminating highly ($r \geq .70$) possessing an item difficulty of 72 percent. This sub-test's item discriminations are, therefore, clearly a function of difficulty. The more difficult the item, the lower its discrimination. A question then arises concerning the 13 items with low discrimination. What is responsible for these items being difficult to the extent that they are not discriminating properly? The possibilities explored in this study are the content (topic) and wording, or structure, of the item.

The reading level of Science was intended to be easy enough so that only the science content of the test would be measured. It is not known if this assertion by the test authors can be applied to the test's administration to hearing impaired students. Since Paragraph Meaning was difficult for the students, there does exist the possibility that the reading level of Science is influencing the performance of

hearing impaired students and thus could be a factor in the test's difficulty.

There is also the matter of the content of the items. In several of the difficult and low discriminating items the length of the stem was, say, six to eight words long. (The lengths of all items in the sub-test range from four to forty-one words.) In these cases of short items and easy wording language difficulty should not be a factor. Rather the possibility of misleading words or an unfamiliar topic arises.

Item No. 17 is the item which discriminated most highly. There is not too much to be said except it appears to be a good item. All of the Upper group and 57 percent of the Lower group correctly answered this item.

17. Many animals change their form as they grow. A toad was first of all –

a **an egg**
b **a cocoon**
c **a moth**
d **a caterpillar**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	100	--	--	--
Lower group	57	11	16	16

Discrimination Coefficient = .70+

Number 36 is an example of a short item with a low language level. Its low discrimination coefficient could possibly be the result of its content, as discussed previously. The distribution of responses suggests that the students may not yet have learned where good soil is hardest to find.

36. Good soil is hardest to find –

a **in the silt of a river bed**
b **on a mountain top**
c **on a grassy plain**
d **in a mountain valley**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	15	24	53	9
Lower group	14	16	35	35

Discrimination Coefficient = .12

The coefficient obtained for item No. 56 shows it to be a very low discriminating item. This item was nearly equally difficult for both groups.

56. A boy filled a balloon with air outdoors on a cold day and sealed it. What happened after he brought the balloon inside his warm house?

a **It got bigger.**
b **It got smaller.**
c **It stayed the same size until some air escaped.**
d **There is not enough information to tell.**

	Percent of Students Choosing Each Alternative			
	a	b	c	d
Upper group	17	49	17	17
Lower group	19	28	28	25

Discrimination Coefficient = –.03

This is an example of an item whose lack of discrimination could lie in its verbal complexity or in its content. The item is testing students' knowledge concerning the density of warm air versus cold air. It would be valuable to know what, if any, difference in performance there would be if the item were re-stated to say, "What happens to the size of a balloon if the air inside it gets warmer?"

Seventy-seven percent of the items in this test had moderate discrimination coefficients, one of them discriminating beyond .70. However, the fact that 23 percent of the items had low discrimination coefficients is sufficient to warrant some concern. Further analyses regarding the language level and curriculum content of this test in particular, and science tests in general, must be made before such tests' validity for hearing impaired students can be adequately determined.

CONCLUSION

The selected sub-tests represent a variety of the subject areas covered in the Intermediate I battery. Some of the sub-tests which were not included are similar to those in this study with respect to item presentation, grade equivalent distributions, overall difficulty, etc. Their degree of discrimination cannot be determined without an item by item analysis. However, projections can be made by comparing these sub-tests with those in the study that display similar psychometric characteristics. This review of

the results of this study will include statements concerning the expected discrimination validity of the sub-tests which were not a part of the study.

The selection of a Paragraph Meaning Sub-test as the standard for battery assignments was made with the awareness that a greater than average degree of homogeneity would result with respect to achievement levels. Therefore, the contrast between the Upper and Lower groups of Paragraph Meaning is not as great as with the remaining sub-tests under study.

The Word Meaning Sub-test can be likened to Paragraph Meaning in that the skills required to do well in these areas are similar. The distribution of grade equivalent scores and item analysis values for both are comparable. Therefore, as with Paragraph Meaning, the items in Word Meaning can be expected to discriminate to a fair degree.

The Social Studies, Part A: Content Sub-test differs from Science only in the specific content of each item. They are presented in the same manner, and students who perform well on one usually do so on the other. Therefore, the items of Social Studies: Content can be expected to discriminate to a similar degree.

Spelling and Arithmetic Computation, although very different in content, produce almost identical group results. Hearing impaired students taking the Intermediate I battery score best on these subtests. Despite this, the items discriminate well. Those two items that did not were located at the beginning or end of the sub-test where a high degree of discrimination is not expected.

The results of Arithmetic Applications probably are a good estimate of the discrimination validity of Arithmetic Concepts. There is a correlation of .77 between student performance on the two, and the average grade equivalent scores differ by two tenths of a year. Both measure the ability to relate numbers to a sense of reasoning. The items of Arithmetic Concepts should be as effective in differentiating levels of achievement as the results of Arithmetic Applications have shown it to be.

The Language Sub-test was not included in this study due to its division into five separate parts. Only Part D of this sub-test probably lacks favorable discrimination validity. It measures Dictionary Skills and is difficult for hearing impaired students.

The Word Study Skills Sub-test, because of its obvious invalidity for hearing impaired students, was not included in this study and cannot be compared to any sub-test that was studied. Since this is a test of phonetics and syllabication, performance here is partly a function of a student's hearing level.

The conclusions reached in this report regarding validity pertain to the population of hearing impaired students who were administered the Intermediate I battery (Form W) as a part of the 1970-71 National Achievement Testing Program. This was one of several studies aimed at obtaining information needed to improve the suitability of an achievement test series for hearing impaired students. Future Achievement Testing Programs of this office will reflect changes which appear necessary as a result of such analyses.

The Reliability of the Stanford Achievement Test, Primary II Battery, for Hearing Impaired Students

Sal DiFrancesca, Ph.D.*

Questions of reliability arise when a test designed for hearing students is used with hearing impaired students. While the reliability of the Stanford Achievement Test has been demonstrated for the standardization population, it remains necessary to do so for the special population of hearing impaired students. The present study evaluates the reliability of the Primary II battery of the 1964 edition of the Stanford Achievement Series. Test-retest methods were used for the two sub-tests which were modified for the hearing impaired (Science and Social Studies Concepts, and Part A of Arithmetic Concepts).[1] The Word Meaning, Paragraph Meaning, Language, Arithmetic Computation, and Arithmetic Concepts, Part B, Sub-tests were evaluated using alternate form methodology, with Forms W and X. The Spelling and Word Study Skills Sub-tests were not evaluated, due to their unsuitability for hearing impaired students as a result of their heavy dependence on auditory skills.

The Primary II level was chosen for study because it was administered to more than 6,500 students, a significant proportion (35 percent) of all students participating in the 1971 National Achievement Testing Program. In addition, it permitted the investigation of the reliability of both modified and unmodified sub-tests.

METHOD

Subjects

The subjects for this study were 192 students selected as follows. First, of all the educational programs voluntarily participating in the 1971 National Testing Program, those were selected which planned to administer the Primary II battery to 30 or more students. From this list of programs, four public residential schools and four public day school programs were selected, geographical representation having been considered. Lists of students to be administered the Primary II battery were then obtained from these eight programs, and their screening test scores[2] were verified so that only students appropriate for testing at the Primary II level were

*At the time the reliability study was conducted, Dr. DiFrancesca was Research Psychologist on the staff of the Office of Demographic Studies.

[1]These modifications are described in detail in Appendix II. Briefly, the modifications consisted of printing into the student's test booklet the text of items intended to be dictated in the original version. Administration involved a combination of teacher dictation with student reading of the printed items.

[2]The screening procedures are described in Appendix II.

retained on the list. Next, subjects meeting any of the following criteria were eliminated from the list:

1. those for whom a nonverbal I.Q. of 75 or less was reported;
2. those below 8 years of age or over 15 years of age;
3. those for whom the Annual Survey did not possess a demographic data file;
4. those who were reported to be mentally retarded, to have perceptual-motor handicaps, or to have a severe visual handicap.

From the students remaining on the lists, 24 students were chosen from each of the eight cooperating programs, for a total subject pool of 192. Eleven subjec s from this final pool were eliminated from the data analyses on the basis of information indicating that they were extremely distracted or had not made a serious attempt to answer the test questions (see the description of the observation procedures in the following section). Finally, any subject who was absent from one or both administrations of a given sub-test was, of course, not included in the data analyses for that sub-test. The final group of subjects ranged in size from 172 to 178 for the various sub-tests.

The average hearing threshold level in the better ear was 86 dB (ISO) for students in this study, with a range from 38 dB to "no response" (120 dB, ISO), and a standard deviation of 17.3 dB. The data on their ages are given in Table 1, along with comparable data for Grade 3 students who were included in the standardization of the Primary II battery.

TABLE 1: CHRONOLOGICAL AGE DATA FOR HEARING IMPAIRED STUDENTS IN THE RELIABILITY STUDY AND FOR GRADE 3 STUDENTS FROM THE STANFORD STANDARDIZATION SAMPLE.

Statistic	Hearing Standardization Sample	Hearing Impaired Reliability Study Sample
Total Students	9,540	189*
Range of Years	8 to 11	8 to 15
Mean Age in Years	8.9	12.1
Standard Deviation Age (Years)	0.5	1.8

*Total subjects selected were 192, but accurate age data were not available for three students.

Observation Procedures

A member of the professional staff of the Annual Survey visited each school in the reliability study as an observer and was present for all testing sessions of the assigned group. The observers were knowledgeable regarding the standard administration procedures specified for the Stanford Primary II battery and were to record student behaviors and test administration errors that in themselves could influence the reliability of the test scores. These observers reported that all the test administrators followed acceptable and standard procedures. It was not necessary to eliminate data from any testing session as a result of inappropriate administration procedures. However, eleven students were described as being distracted and making no serious attempt to answer the test questions. As a result, data from these 11 students were excluded from the analyses.

Testing Procedures

The selection of the test administrators was left to the participating schools with the stipulation that they be typical of individuals who usually administered tests in that school. Each testing group had a different test administrator who remained with the group for all test sessions.

Within each participating school the subjects were divided into two groups, labeled A and B. All such testing groups were selected to be as similar as possible on variables such as age, degree of hearing loss, and I.Q. scores, both within individual schools and among all the participating programs. Students in "A" groups were tested with Form W first, followed by Form X. In the "B" groups the order of testing was reversed, Form X being administered first. The two modified sub-tests (Science and Social Studies Concepts, and Arithmetic Concepts, Part A) were administered in exactly the same form on both occasions. Each testing group received two sub-tests per day for three consecutive days. Re-testing occurred within a two to four week period.

RESULTS

The raw score means and standard deviations and the average grade equivalents for the six sub-tests of both test forms for groups A and B are presented in Table 2.

TABLE 2: COMPARISON OF SCORES ON FIRST TEST AND RETEST OF PRIMARY II BATTERY, STANFORD ACHIEVEMENT TEST RELIABILITY STUDY, HEARING IMPAIRED STUDENTS.

GROUP A		FIRST TEST–FORM W			RETEST–FORM X		
Primary II Battery Sub-tests	Number of Students	Average Raw Score	Standard Deviation (Raw Score)	Average Grade Equivalent	Average Raw Score	Standard Deviation (Raw Score)	Average Grade Equivalent
Word Meaning	89	15.6	4.0	2.7	18.4	3.9	2.8
Paragraph Meaning	89	31.6	7.5	2.9	31.3	8.1	2.9
Science & Social Studies*	88	15.3	4.8	2.2	15.0	4.8	2.2
Language**	90	43.0	8.6	3.6	41.0	9.7	3.1
Arithmetic Computation	88	40.6	16.2	4.1	41.4	16.2	4.1
Arithmetic Concepts***	86	21.0	8.6	3.3	22.7	8.5	3.1
GROUP B		**FIRST TEST–FORM X**			**RETEST–FORM W**		
Primary II Battery Sub-tests	Number of Students	Average Raw Score	Standard Deviation (Raw Score)	Average Grade Equivalent	Average Raw Score	Standard Deviation (Raw Score)	Average Grade Equivalent
Word Meaning	84	18.2	4.3	2.8	15.6	4.5	2.7
Paragraph Meaning	83	32.5	8.0	3.0	31.3	8.1	2.9
Science & Social Studies*	90	14.7	4.8	2.2	15.4	4.7	2.2
Language**	88	40.5	9.7	3.1	42.8	10.1	3.6
Arithmetic Computation	87	39.0	15.4	3.9	40.6	15.2	4.1
Arithmetic Concepts***	87	20.9	9.0	2.9	21.7	9.2	3.2

*Form W-HI of this sub-test was administered at both first and second testing.

**The difference between the mean grade equivalent scores for Forms W and X of this sub-test is significant beyond the .01 level, by t-test for the difference between sample means.

***Part A of this sub-test was administered in the Form W-HI at both test sessions. Part B was varied between Form W and Form X as indicated in the table headings.

Inspection of the data in Table 2 indicates that the differences in mean grade equivalent scores between Form X and Form W were trivial except for the Language Sub-test, where a difference of 0.5 grade equivalent was observed. Significance testing with the t-test for the difference between two sample means indicates that all differences are non-significant, except for the difference in the Language Sub-test scores, which was significant beyond the .01 level.

The reliability coefficients obtained are presented in Table 3. All were obtained by using Pearson product-moment correlations between the raw scores obtained by students on the two forms for groups A and B combined. Since the Science and Social Studies Concepts Sub-test was administered in exactly the same form on the two occasions, the reported

TABLE 3: RELIABILITY COEFFICIENTS FOR SUB-TESTS OF PRIMARY II BATTERY, STANFORD ACHIEVEMENT TEST, GROUPS A AND B COMBINED.

Primary II Battery Sub-tests	Number of Students	Reliability Coefficients
Word Meaning	173	.69*
Paragraph Meaning	172	.82*
Science & Social Studies	178	.75**
Language	178	.81*
Arithmetic Computation	175	.95*
Arithmetic Concepts	173	.91*

*Coefficient of stability and equivalence
**Coefficient of stability

coefficient is properly called a coefficient of stability of scores over a two to four week period. The coefficients for the other sub-tests are properly considered coefficients of stability and equivalence, since they were influenced both by the effects of the passage of time and by the specific differences in content between the two different forms of the test employed. All the reported coefficients are significantly different from zero beyond the .001 level, by means of the t-test for the significance of a correlation coefficient.

The standard errors of measurement (SE_m) for the various sub-tests are presented in Table 4 in raw score form. These standard errors of measurement are listed for the hearing impaired reliability study group and for the Grade 2 and Grade 3 students who made up the standard.zation group for the Primary II battery. The chances are approximately 2:1 that a student's "true score" (the mean of all obtained scores from a very large number of administrations of the test) does not differ from the obtained score by more than one standard error of measurement. The chances are about 19:1 that the obtained score does not differ from the "true score" by more than two standard errors of measurement. The SE_m is therefore another way to express test reliability which is especially useful in the case of an individual student's scores.

These raw score SE_m's correspond to differences of from 0.2 to 0.7 grade equivalents (i.e., 2 to 7 months), depending on the particular subtest and the position within that test's range at which the student's obtained score falls.

DISCUSSION

The reliability coefficients reported range from adequate to good for the various sub-tests. Direct comparability with the results for the standardization group is not possible since the reliabilities reported for the standardization group are of the split-half and Kuder-Richardson Formula 20 varieties, rather than the alternate-form and test-retest varieties used here. The figures reported for the Arithmetic Computation and Arithmetic Concepts Sub-tests (.95 and .91, respectively) are quite satisfactory and indicate a high degree of comparability between Forms W and X for hearing impaired students in these subject areas. For the other sub-tests evaluated, the range of coefficients from .69 to .82 indicates substantial degrees of commonality between the two test forms, but also points to more than trivial amounts of difference in item content. The one sub-test evaluated by test-retest methods (Science and Social Studies Concepts) showed an acceptable coefficient (.75), but one that might have been expected to be higher given the relatively short test-retest interval. Some remembering of specific items or item responses should have occurred, a situation which tends to increase reliability coefficients.

While the obtained grade equivalents on the Language Sub-test differed significantly between the two forms, the reliability coefficient remained acceptably high, indicating that the rank ordering of students on the two forms was fairly well preserved. A student who ranked tenth in his score on the Form W version of the test would tend to rank very near

TABLE 4: RAW SCORE STANDARD ERRORS OF MEASUREMENT FOR SUB-TESTS OF THE PRIMARY II BATTERY, HEARING IMPAIRED STUDENTS AND STANDARDIZATION GROUP STUDENTS.

	Standard Error of Measurement in Raw Score Units		
Sub-test Name	Hearing Impaired Reliability Study	Standardization Grade 2	Standardization Grade 3
Word Meaning	2.4	2.5	2.2
Paragraph Meaning	3.3	3.2	2.8
Science & Social Studies Concepts	2.4	2.8	2.6
Language	4.1	4.0	3.6
Arithmetic Computation	3.5	2.1	2.6
Arithmetic Concepts	2.7	2.9	2.8

tenth on Form X as well. The observed differences in grade equivalents on the Language Sub-test are apparently due to specifics of content, which made Form X significantly more difficult for the hearing impaired students. It is not known at this point just what these specifics were.

The comparable reliability figures for other forms of the Stanford test and for the other battery levels of the series are not known for hearing impaired students. Because of the similarity of item format and the distribution of item difficulties in the other batteries, the reliability figures are probably comparable to those obtained in this study of the Primary II battery. Nevertheless, in the absence of the data, this remains only an informed guess.

In short, the reliability coefficients obtained for hearing impaired students range from adequate to good for the six sub-tests which were evaluated by alternate form and test-retest methodology on the Primary II battery. These results suggest that results obtained from either Form W or Form X of the test are quite comparable, with the sole exception of the Language Sub-test, which appears to be significantly more difficult for hearing impaired students in Form X than in Form W.

APPENDIX I

DESCRIPTION, QUALIFICATIONS, AND LIMITATIONS OF THE ACHIEVEMENT TESTING PROGRAM

The Office of Demographic Studies first became involved in the area of achievement testing in the spring of 1969. One of the important areas in which data were needed, according to the project's National Advisory Committee, was that of the outcomes of the educational process as measured by achievement tests. Information subsequently obtained from educational programs for the hearing impaired indicated that the Stanford Achievement Test was the most widely used measure of academic achievement for hearing impaired students. Consequently, the Stanford Series was used in a national survey of achievement testing of hearing impaired students conducted during the spring of 1969. The results of the more than 12,000 tests administered during that first program confirmed the general knowledge of educators of the deaf that hearing impaired students score substantially below average levels of achievement attained by their hearing agemates.[1] Of greater importance was the fact that the results indicated that large numbers of the students tested were receiving test batteries too advanced for their achievement level, with the result that many scores were at or below the level where guessing or random response becomes a major determinant of the obtained scores. It was therefore necessary to revise the measuring instruments themselves before an adequate measurement of the achievement of hearing impaired students could be obtained.

[1] The results of this program have been presented in detail in publications D-1 and D-2 from the Office of Demographic Studies, listed on the inside back cover.

A second National Testing Program was undertaken in the spring of 1971, incorporating three innovations designed to handle the technical measurement problems encountered two years earlier. These were (1) a screening test procedure to determine the appropriate level at which a student should be tested; (2) practice tests to familiarize students with the mechanics of test-taking, a set of skills which seemed to be deficient in many of the students tested in 1969; and (3) a modification of the sub-tests which in the original version of the test were intended to be dictated by the teacher. The dictated items were, instead, printed in the student's test booklet so that the benefits of both dictation and reading of the item were available to the hearing impaired student. It is upon the data from this second National Testing Program that the studies in the present report have been based.

DESCRIPTION OF THE STANFORD ACHIEVEMENT TESTS

The Stanford test is described by its authors as:

> . . . comprehensive achievement tests developed to measure the important knowledges, skills, and understandings commonly accepted as desirable outcomes of the major branches of the elementary curriculum. The tests are intended to provide dependable measures of these outcomes, comparable from subject to subject and grade to grade,

for use in connnection with improvement of instruction, pupil guidance, and evaluation of progress.[1]

The 1971 National Testing Program used Form W of the 1964 edition of the Stanford test for the overall testing. Form X of the 1964 edition was used for the screening test procedure and for the retesting in the reliability study described in this publication. The five batteries of this 1964 edition (Primary I, Primary II, Intermediate I, Intermediate II, and Advanced) each cover academic materials in various subject areas appropriate for students within a specific grade range. For example, the Primary I battery is intended for use from the middle of Grade 1 to the middle of Grade 2. The titles of the sub-tests included in each battery are listed in Table A.

[1] Kelley, T. L., Madden, R., Gardner, E. F., and Rudman, H.C. *Stanford Achievement Test: Directions for Administering Primary I Battery.* New York: Harcourt, Brace & World, Inc., 1965, p. 2.

The content of the Stanford test was based on a survey of the materials typically included in school curricula at the various grade levels included in the test, and the standardization of the final form of the test was carried out in school systems across the country. Curricula for the hearing impaired were not surveyed, and educational programs for the hearing impaired were not included in the standardization procedures.

PARTICIPANTS IN THE 1971 NATIONAL TESTING PROGRAM

All educational programs for the hearing impaired known to the Annual Survey in fall of 1970, 776 programs enrolling approximately 48,000 students, were contacted by letter and invited to participate in the testing program. Test materials and scoring services were offered free of charge to the participating programs. A total of 292 programs accepted the

TABLE A: SUB-TESTS CONTAINED IN SUCCESSIVE BATTERY LEVELS OF THE STANFORD ACHIEVEMENT TEST SERIES, FORM W.

Primary I	Primary II	Intermediate I	Intermediate II	Advanced
Word Reading	Word Meaning	Word Meaning	Word Meaning	
Parag. Meaning	Parag. Meaning	Parag. Meaning	Parag. Meaning	Parag. Meaning
Vocabulary	Science & Social Studies Concepts			
Spelling	Spelling	Spelling	Spelling	Spelling
Word Study Skills	Word Study Skills	Word Study Skills		
	Language	Language	Language	Language
	Arithmetic Computation	Arithmetic Computation	Arithmetic Computation	Arithmetic Computation
Arithmetic	Arithmetic Concepts	Arithmetic Concepts	Arithmetic Concepts	Arithmetic Concepts
		Arithmetic Applications	Arithmetic Applications	Arithmetic Applications
		Social Studies	Social Studies	Social Studies
		Science	Science	Science

invitation and tested over 19,000 students. Participation was voluntary on the part of the educational programs, and no follow-up effort was employed to encourage non-respondents or non-participating programs to participate in the national testing. The reason most frequently cited for non-participation was that the given program enrolled only pre school students or other students too young to be tested. Other programs declined participation because they had insufficient staff to administer the tests, because they were complying with school district testing programs using tests other than the Stanford, because they were itinerant programs with small numbers of students scattered across wide geographic areas, or because of reservations about administering the Stanford test to hearing impaired students. The final number of students who were tested at each battery level is reported in Table B.

TABLE B: NUMBER OF ACHIEVEMENT TESTS ADMINISTERED ACCORDING TO TEST BATTERY LEVEL: SPRING 1971.

Test Battery Level	Number	Percent
All Levels	19,037	100.0
Primary I	6,786	35.6
Primary II	6,655	35.0
Intermediate I	3,215	16.9
Intermediate II	1,566	8.2
Advanced	815	4.3

QUALIFICATIONS AND LIMITATIONS OF THE TESTING RESULTS

Many of the qualifications and limitations of the testing results have been well stated in previous publications from the Office of Demographic Studies dealing with the 1971 testing program,[1] and the reader is urged to review these statements before making use of the information contained in this report.

Of major concern is the non-random method of selection of students for inclusion in the testing program. Appendix III to Office of Demographic Studies publication D-9 reviews the demographic characteristics of the students tested in the 1971 program in relation to the characteristics of all students for whom information is available to the Annual Survey. Examination of these data suggests that the differences are not great except in the case of the age distribution. This is an expected and acceptable deviation, since preschool students and students under age 8 (who constitute a substantial proportion of the total Annual Survey group) are not ordinarily tested with achievement tests of the Stanford type.

Another characteristic of the age distribution is noteworthy, namely, the age distributions of students taking particular battery levels. The age range of hearing impaired students receiving a given battery is substantially broader than that encountered among hearing students. For example, 155 students age 18 or above received primary level test batteries on the basis of screening test scores indicating that their achievement levels are within the primary range. Regardless of the accuracy of such assignments of testing levels, the presumed disparity between the content matter of primary level tests and the typical interest patterns of 18-year old students introduces a source of extraneous variance unrelated to actual academic ability but related instead to the motivational characteristics of older students being asked to respond to such materials. This is a measurement problem which will need to be handled in some other manner in future testing programs. The age distributions of hearing impaired students receiving each of the five batteries are presented in Table C.

The reports contained in this and previous publications regarding the 1971 Achievement Testing Program have presented data relating to the item analysis figures, means and standard deviations of grade equivalents obtained by hearing impaired students according to degree of hearing loss, patterns of differential performance between the hearing impaired and standardization groups, reliability, and discriminative validity of the tests used in the 1971 program. Future reports will concentrate on the influence of other demographic characteristics on the achievement levels of hearing impaired students, on the influence of guessing or random response on the test scores, and on the intercorrelations among the sub-tests of each achievement battery. The achievement testing activities of the Office of Demographic Studies continue to be directed to the goal of improving the accuracy and usefulness of the tests used in the field of education of hearing impaired children and youth and to the assessment of the outcomes of the educational process by means of such improved tests.

[1] See Office of Demographic Studies publication D-8, pp. 2-3, and publication D-9, pp. 6-7.

TABLE C: NUMBER OF STANFORD ACHIEVEMENT TESTS INCLUDED IN THIS REPORT BY AGE AND TEST BATTERY LEVEL: SPRING 1971.

Age	Battery Level					
	All Levels	Primary I	Primary II	Intermediate I	Intermediate II	Advanced
All Ages Tested	19,037	6,786	6,655	3,215	1,566	815
Unknown Age	2,129	800	722	394	125	88
Total Students Included in This Report	16,908	5,986	5,933	2,821	1,441	727
Under 6	31	27	2	1*	1*	–
6	341	335	3	2	1*	–
7	454	425	25	2	1*	1*
8	697	606	85	6	–	–
9	971	718	238	11	3	1*
10	1,297	779	466	42	9	1*
11	1,600	765	737	72	25	1*
12	2,316	913	1,092	232	66	13
13	1,547	401	758	288	84	16
14	1,573	352	674	392	117	38
15	1,455	230	538	409	215	63
16	1,315	161	451	355	210	138
17	1,244	119	361	375	239	150
18	1,154	108	300	325	248	173
19	639	35	142	208	152	102
20	222	10	47	86	54	25
21 & Over	52	2	14	15	16	5

*These results are highly improbable and may reflect an error in scoring or the age given for the student who took the test.

APPENDIX II

STANDARDIZED TESTING PROCEDURES DEVELOPED FOR THE SPRING 1971 ACHIEVEMENT TESTING PROGRAM

The analyses of data collected from the 1969 Achievement Testing Program indicated that different methods of administering the tests were being used among individual school and class programs. As test scores can be affected by the manner in which the test is given, it became necessary to establish uniform testing procedures. This served the purpose of making test administration procedures consistent throughout the schools and classes participating in the spring, 1971, testing program. It also ensured that test scores would be comparable from teacher to teacher and school to school. A description of the standardized procedures implemented to collect the data in this report is given below.

SCREENING TESTING OR PRE-TESTING OF STUDENTS

Analyses of the 1969 testing data demonstrated that many sub-tests, particularly at the Intermediate and Advanced battery levels, were not showing true differences between good and poor students. This occurred mainly because students were receiving test battery levels too high or too difficult for them. The number of items they were able to answer correctly was insufficient to show actual achievement differences, and scores tended to cluster about a chance or guessing range.

This guessing factor may result in a student's score being spuriously affected by the test battery level he receives. Generally, by guessing alone, the higher the battery level administered, the higher will be the scores. For example, if a beginning first grade student were administered the Social Studies Sub-test of the Advanced battery and merely guessed at each question, he would likely receive a 4.6 Grade Equivalent score. The criteria used to select test battery levels for students varied throughout the country. It was therefore necessary to establish valid battery selection methods that would be consistent among the participating schools. A screening testing procedure was implemented to accomplish this goal.

The selection of the screening test was, for the most part, based on the internal analyses of the 12,000 achievement records collected two years earlier. The search was to find one sub-test within the Stanford Series which best indicated how well a student would perform on the remainder of the sub-tests in the full battery. On the basis of various statistical analyses, the Paragraph Meaning Sub-test consistently proved to be the best predictor of overall student performance. In setting the specifications for using a Paragraph Meaning score to select the full battery, statistical adjustments were made which allowed for the fact that younger students generally scored higher in reading than on other test content areas, while older students scored relatively lower on Reading than in the remainder of the test, e.g., Arithmetic Computation.

Two levels of screening tests were used, one appropriate for students achieving at a general level of the end of the fourth grade and below, and one for students estimated to be functioning at the beginning fifth grade level and above. In ordering screening test materials, the participating programs were asked to estimate the number of students maintaining a general academic level within each of these broad categories. The Paragraph Meaning Sub-tests from the Primary II and Intermediate II batteries of the Stanford Series, Form X, were employed as the screening instruments. The school administered and scored the screening test. The number of items the student answered correctly was used to select the most valid battery level for him. Guidelines for using screening test scores were formulated and set by the Annual Survey.

PRACTICE TESTING TO INSTRUCT STUDENTS IN TEST TAKING PROCEDURES

The directions to administer parts of the Stanford test and the question-answer format of some test items proved difficult to follow for many students in special programs for the hearing impaired. They lacked exposure to this type of testing procedure. In analyzing the 1969 achievement test results, it became clear that many students took the test not understanding the test structure or how to mark their answers. The seriousness of this problem led the Annual Survey to develop sets of practice tests appropriate to each battery level of the Stanford Series. Samples of the test directions, questions, and answer marking procedures were included in the practice materials, along with an explanation of the purposes of academic achievement testing. Teacher manuals were developed to accompany the practice tests.

Participating programs received a practice test for each student and were requested to administer them two to four days prior to the Stanford full batteries. The practice tests were to be used directly to teach test-taking mechanics to the students and prepare them for their best performance on the real test. As the teachers gave the practice session, they too became better prepared for administering the Stanford test.

SPECIAL EDITION OF THE PRIMARY LEVEL TEST BATTERIES

The Primary I and II test levels, those intended for the academic range of the middle of Grade 1 to the end of Grade 3, contain many sub-tests structured to be administered by oral presentation. A hearing impaired student's response to a dictated question may be a function of his receptive communication skill and not his knowledge of the answer. Previously, schools attempted various procedures to overcome this problem – overheads, blackboards, and the like. To standardize presentation of the dictated sub-tests and make their design more valid, the Annual Survey arranged with the test authors and publisher for a special edition. This applied to the Primary I and II levels only and was called Form W-HI. The Intermediate and Advanced test levels are self-administering and contain no dictated test questions.

Within the Form W-HI edition, those test questions previously to be strictly dictated were also printed in the test booklet itself. The teacher was to dictate the question and then direct the student to read it in his own booklet before marking his answer. The procedure served to make uniform the administration of dictated sub-tests. In the Primary I, Form W-HI, modifications were made in the Vocabulary and Arithmetic Sub-tests. The Science and Social Studies Concepts and Arithmetic Concepts Sub-tests were modified at the Primary II level, Form W-HI.

APPENDIX III

SCHOOLS AND CLASSES THAT PARTICIPATED IN THE ACHIEVEMENT TESTING PROGRAM

ALABAMA
- Alabama Institute for the Deaf and Blind
- Blossomwood Elementary School
- Children's Center of Montgomery, Inc.
- Holt Elementary School
- University of Montevallo Speech and Hearing Clinic

ALASKA
- Anchorage Borough School District

ARIZONA
- Arizona State School for the Deaf and Blind
- Phoenix Day School for the Deaf

ARKANSAS
- Jenkins Memorial Children's Center

CALIFORNIA
- Alhambra City School District
- Anaheim Union High School District
- Bellflower Unified School District
- Mary E. Bennett School for the Deaf
- Butte County Schools
- California School for the Deaf, Riverside
- Cedarcreek School for the Deaf
- Centralia School District
- Chula Vista City School District
- Covina Valley Unified School District
- El Centro Elementary School District
- Escondido Union School District
- Garden Grove Unified School District
- Goleta Union Elementary School District
- Kern County Schools
- La Mesa-Spring Valley School District
- Lancaster Elementary School District
- Livermore Valley Joint Unified School District
- Lompoc Unified School District
- Marin County Schools
- Marlton Elementary School
- Monterey County Schools
- Mt. Diablo Unified School District
- Oakland City Unified School District
- Orange Unified School District
- Pasadena City Unified School District
- Placer County Public Schools
- Riverside Unified School District
- San Bernardino County Schools
- San Diego Unified School District
- San Francisco County Schools
- San Jose City Unified School District
- San Juan Unified School District
- Santa Ana Unified School District
- Santa Clara Unified School District
- Santa Rosa City School District
- Simi Valley Unified School District
- Solano County Schools
- Stockton Unified School District
- Sutter County Schools
- Tehama County Public Schools
- Tulare County Schools
- Tulare Union High School District

COLORADO
- Colorado School for the Deaf and Blind
- John Evans School
- Meadow Elementary School

CONNECTICUT
- American School for the Deaf
- Class for Preschool Hearing Impaired Children, Hartford
- East Hartford Public Schools
- Green Acres School
- Hamden-New Haven Cooperative Educational Center
- Magrath School
- Mystic Oral School for the Deaf
- West Haven Department of Special Education

DELAWARE
- Margaret S. Sterck School for Hearing Impaired

DISTRICT OF COLUMBIA
- Capitol Region Model Secondary School (MSSD)
- Kendall School for the Deaf
- Speech and Hearing Center--Public Schools of the District of Columbia

FLORIDA
- Brevard County Public Schools
- Florida School for the Deaf and Blind
- Leon County Program for Hearing Impaired Children
- Robert McCord Oral School
- Palm Beach County Schools
- Rock Lake Elementary School

GEORGIA
- Atlanta Public Schools
- Atlanta Speech School, Inc.
- Cobb County Board of Education
- Lawton B. Evans School
- Houston Speech and Hearing School
- Robert Shaw Center

HAWAII
- Central Intermediate School
- Diamond Head School for the Deaf
- McKinley High School

IDAHO
- Idaho School for the Deaf and Blind

ILLINOIS
- Bell Elementary School
- Bi-County Oral Deaf Program
- Black Hawk Hearing Handicapped Program
- Champaign Community Schools
- Chicago Vocational High School
- Decatur Public School District
- Elim Christian School for the Exceptional Child
- Ericson School
- Illinois School for the Deaf
- Jamieson School
- Marquette Elementary School
- Thomas Metcalf School
- Morrill Elementary School
- Northern Suburban Special Education District
- Northwest Suburban Special Education Organization
- Northwestern Illinois Association
- Perry School
- Ray School
- Reinberg School
- Scammon School
- Shields Elementary School
- South Metropolitan Association for Low-Incidence Handicapped
- Special Education District of Lake County
- Springfield Public Schools
- West Suburban Association for the Hearing Handicapped
- James Ward Elementary School

INDIANA
- Glenwood Elementary School
- Hammond Public Schools
- Indiana School for the Deaf
- Marion Community Schools
- Morrison-Mock School
- Fayette County Schools Corporation

IOWA
- Black Hawk-Buchanan County Board of Education
- Cedar Rapids Community Schools
- Hope Haven School
- Iowa School for the Deaf
- Wilson School-Oral Deaf Department

KANSAS
- Kansas School for the Deaf
- Wichita Public Schools

KENTUCKY
- Kentucky School for the Deaf
- Louisville Public Schools

LOUISIANA
- Acadia Parish School Board
- Lafayette Parish School Board
- Louisiana School for the Deaf
- Monroe City Schools
- Sunset Acres School

MAINE
- Governor Baxter State School for the Deaf

MARYLAND
- Baltimore County Department of Special Education
- Maryland School for the Deaf
- Montgomery County Public Schools
- Prince George's County Public Schools

MASSACHUSETTS
- Belmont Public Schools
- Beverly School for the Deaf
- Boston School for the Deaf

Peter Bulkeley School
Clarke School for the Deaf
Lawrence Primary Program for the Deaf
Leominster Day Classes for the Hearing Impaired
Horace Mann School for the Deaf
Mercer School
Willie Ross School for the Deaf
Upsala Street School
Woburn Day Class Program
Worcester County Hearing and Speech Center

MICHIGAN
Howard D. Crull Intermediate School (Roosevelt Elementary)
Detroit Day School for Deaf
Douglas School
Durant-Tuuri-Mott School
Escanaba Area Jr. High School
Ferndale Public Schools
Handley School
Ida Public Schools
Kalamazoo Public Schools
Ann J. Kellogg School
Lakeview Elementary School
Lakeview Public Schools
Lindemann Elementary School
Lutheran School for the Deaf
Marquette Elementary
Michigan School for the Deaf
Oakland Schools
Public School Program for Deaf and Hard-of-Hearing, Jackson
Traverse City Public Schools
Utica Schools

MINNESOTA
Duluth Public Schools
Minnesota School for the Deaf
St. Paul Area Program for Impaired Hearing

MISSISSIPPI
Mississippi School for the Deaf
Popp's Ferry Elementary School

MISSOURI
Central Institute for the Deaf
Delaware Elementary School
Litzsinger School
Missouri School for the Deaf
St. Louis County Special School District for the Handicapped
School District of Kansas City

MONTANA
Montana State School for the Deaf and Blind

NEBRASKA
Nebraska School for the Deaf
Omaha Public Schools
Prescott Acoustically Handicapped Unit

NEVADA
Ruby S. Thomas Elementary School

NEW HAMPSHIRE
Crotched Mountain School for the Deaf

NEW JERSEY
Bruce Street School
Class for the Hard of Hearing, Kearny
Cumberland County Public Schools
Hackensack Program for the Deaf
Marie H. Katzenbach School for the Deaf
Millburn Avenue School
Township Public Schools, Neptune
Woodbridge Public School System

NEW MEXICO
New Mexico School for the Deaf

NEW YORK
Board of Cooperative Educational Services, Nassau
Board of Cooperative Educational Services of Washington, Warren and Hamilton Counties
Board of Cooperative Educational Services, Suffolk County II
Board of Cooperative Educational Services, Suffolk County III
Catholic Charities Day Classes for Deaf Children
Mill Neck Manor Lutheran School
New York School for the Deaf–White Plains
New York State School for the Deaf–Rome
Rochester School for the Deaf
St. Francis De Sales School for the Deaf
St. Joseph's School for the Deaf
St. Mary's School for the Deaf
School for Language and Hearing Impaired Children–Public School 158
Union-Endicott Central School District

NORTH CAROLINA
Eastern North Carolina School for the Deaf
North Carolina School for the Deaf

NORTH DAKOTA
Longfellow School
North Dakota School for the Deaf

OHIO
- Alexander Graham Bell School for the Deaf, Cleveland
- Canton Public Schools
- Kennedy School for the Deaf
- Kent Public Schools
- Lakewood Public Schools
- Lorain Board of Education
- Mansfield City Schools
- Ohio School for the Deaf
- Program for Physically Handicapped, Toledo
- Springfield City Schools
- Youngstown Public Schools
- Zanesville Classes for Deaf

OKLAHOMA
- Kerr Junior High School
- Oklahoma City Public Schools
- Oklahoma School for the Deaf
- University of Oklahoma Medical Center

OREGON
- Oregon State School for the Deaf
- Portland Public Schools
- Tucker-Maxon Oral School
- Washington County Intermediate Education District

PENNSYLVANIA
- DePaul Institute
- Ebensburg State School and Hospital
- Erie City School District
- Home of the Merciful Saviour for Crippled Children
- Willis and Elizabeth Martin School
- Pennsylvania School for the Deaf
- Pennsylvania State Oral School for the Deaf
- Programs for Speech and Hearing Handicapped:
  - Centre County Schools
  - Clinton County Schools
  - Fayette County Schools
  - Northampton County Schools
- Western Pennsylvania School for the Deaf

RHODE ISLAND
- Rhode Island School for the Deaf

SOUTH CAROLINA
- Florence County School District #3
- Pate Elementary School
- South Carolina School for the Deaf and Blind

SOUTH DAKOTA
- South Dakota School for the Deaf

TENNESSEE
- Knox County Public Schools
- Memphis Parents' School for Deaf and Aphasic
- Tennessee School for the Deaf

TEXAS
- Abilene Public Schools–Day Class for the Deaf
- Austin Independent School District
- Bexar County School for the Deaf
- P.F. Brown Elementary School
- The Callier Hearing and Speech Center
- Corpus Christi Independent School District
- County-Wide Area Day School, El Paso
- Dallas Independent School District
- Hereford Independent School District
- Houston Independent School District
- Houston School for Deaf Children
- Tarrant County Day School for Deaf
- Texas School for the Deaf
- Wichita Falls Independent School District

UTAH
- Utah Schools for the Deaf and Blind
- Utah State University–Edith Bowen Laboratory School

VERMONT
- Austine School for the Deaf

VIRGINIA
- Arlington County Public Schools
- Charlottesville Public Schools
- Diagnostic, Adjustive and Corrective Center for Learning
- Virginia School for the Deaf and Blind
- Virginia State School for the Deaf at Hampton

WASHINGTON
- Bellevue Public Schools
- Bellingham School District #501
- Edna E. Davis School
- Northshore School District #417
- Seattle Public Schools
- Shoreline School District #412
- Washington State School for the Deaf

WEST VIRGINIA
- West Virginia School for the Deaf and the Blind

WISCONSIN
- City District Public Schools, La Crosse
- Day School for the Deaf, Wausau
- Lincoln Elementary, Eau Claire
- Madison Public Schools
- Pleasant Hill School
- St. John's School for the Deaf
- School for the Deaf, Green Bay
- School for the Deaf, Oshkosh
- E. H. Wadewitz School
- Wisconsin School for the Deaf

WYOMING
- Wyoming School for the Deaf

REPORTS FROM THE ANNUAL SURVEY OF HEARING IMPAIRED CHILDREN AND YOUTH

SERIES D

No. 1 Academic Achievement Test Performance of Hearing Impaired Students — United States: Spring 1969

No. 2 Item Analysis of Academic Achievement Tests Hearing Impaired Students — United States: Spring 1969

No. 3 Additional Handicapping Conditions, Age at Onset of Hearing Loss, and Other Characteristics of Hearing Impaired Students — United States: 1968-69

No. 4 Type and Size of Educational Programs Attended By Hearing Impaired Students — United States: 1968-69

No. 5 Summary of Selected Characteristics of Hearing Impaired Students — United States: 1969-70

No. 6 Audiological Examinations of Hearing Impaired Students — United States: 1969-70

No. 7 Characteristics of Hearing Impaired Students Under Six Years of Age — United States: 1969-70

No. 8 Item Analysis of an Achievement Testing Program for Hearing Impaired Students — United States: Spring 1971

No. 9 Academic Achievement Test Results of a National Testing Program for Hearing Impaired Students — United States: Spring 1971

No. 10 Characteristics of Hearing Impaired Students by Hearing Status — United States: 1970-71

No. 11 Studies in Achievement Testing, Hearing Impaired Students — United States: Spring 1971

SPECIAL REPORTS FROM THE OFFICE OF DEMOGRAPHIC STUDIES, GALLAUDET COLLEGE

SERIES C

No. 1 National Survey of State Identification Audiometry Programs and Special Educational Services for Hearing Impaired Children and Youth — United States: 1972